

168:1255-1270. It encodes the first 8 and a half amino terminal SCRs of CRI. The invention concerns analogues of "regulator of complement activation" proteins or truncated, hybrid or recombinant forms of them. CRI-4 is a preferred truncated form and a number of specified substitution variants of it are claimed in which certain positions in SCR-1 which have been identified as important for the degree of C3b- and C4b-binding are substituted by amino acids from the corresponding positions in SCR-8. The specification does not contain the CRI-4 sequence; the sequence given here was constructed from the full-length CRI amino acid sequence having GENESB0 accession number R11810 and descriptions in the disclosure.

Sequence 343 AA;

Query Match 100.0%; Score 186; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 1.5e-18;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVFELVGPSTYCTSDNDQVIGMSG 33  
|||||  
DB 154 CNPSSGGRKVFELVGPSTYCTSDNDQVIGMSG 186

RESULT 8  
R28547  
ID R28547 standard; peptide: 543 AA.  
AC R28547;  
DT 19-MAR-1993 (first entry)  
DE CRI-4 (52S, 53S, 54P) analogue.  
KW short consensus repeat; regulator of complement activation;  
KM C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.  
FH Key  
FT region 1.60  
FT /label= SCR-1  
FT 61..122  
FT /label= SCR-2  
FT 451..510  
FT /label= SCR-8  
FT 511..543  
FT /label= SCR-9  
FT /note="TRUNCATED"  
FT misc\_difference 52  
FT /note="Thr substituted by Ser (SCR-8)"  
FT misc\_difference 53  
FT /note="Gly substituted by Ser (SCR-8)"  
FT misc\_difference 54  
FT /note="Ala substituted by Pro (SCR-8)"  
FN EP-512733-A.  
PD 11-NOV-1992.  
PF 28-APR-1992; 303826.  
PR 03-MAY-1991; US-695514.  
PA (UNIV ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Kiyoch M;  
DR WPI; 92-375009/46.  
PT Complement activily regulator protein analogues - useful for  
PT treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11: Fig 2 and R11810: 23pp; English.  
CC The CDNA clone designated CRI-4 was described in J.Exp.Med.(1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCRs of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-1 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-8. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESB0  
CC accession number R11810 and descriptions in the disclosure.  
SQ Sequence 543 AA;

Query Match 100.0%; Score 186; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 1.5e-18;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVFELVGPSTYCTSDNDQVIGMSG 33  
|||||  
DB 154 CNPSSGGRKVFELVGPSTYCTSDNDQVIGMSG 186

RESULT 9  
R28548  
ID R28548 standard; peptide: 543 AA.  
AC R28548;  
DT 19-MAR-1993 (first entry)  
DE CRI-4 (57V, 59K) analogue.  
KW short consensus repeat; regulator of complement activation;  
KM C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.  
FH Key  
FT region 1.60  
FT /label= SCR-1  
FT 61..122  
FT /label= SCR-2  
FT 451..510  
FT /label= SCR-8  
FT 511..543  
FT /label= SCR-9  
FT /note="TRUNCATED"  
FT misc\_difference 57  
FT /note="Arg substituted by Val (SCR-8)"  
FT misc\_difference 59  
FT /note="Arg substituted by Lys (SCR-8)"

FN EP-512733-A.  
PD 11-NOV-1992.  
PF 28-APR-1992; 303826.  
PR 03-MAY-1991; US-695514.  
PA (UNIV ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Kiyoch M;  
DR WPI; 92-375009/46.  
PT Complement activily regulator protein analogues - useful for  
PT treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11: Fig 2 and R11810: 23pp; English.  
CC The CDNA clone designated CRI-4 was described in J.Exp.Med.(1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCRs of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-1 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-8. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESB0  
CC accession number R11810 and descriptions in the disclosure.  
SQ Sequence 543 AA;

Query Match 100.0%; Score 186; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 1.5e-18;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVFELVGPSTYCTSDNDQVIGMSG 33  
|||||  
DB 154 CNPSSGGRKVFELVGPSTYCTSDNDQVIGMSG 186

RESULT 10  
R28549  
ID R28549 standard; peptide: 543 AA.  
AC R28549;  
DT 19-MAR-1993 (first entry)

```

R28544
ID R28544 standard; peptide: 543 AA.
AC R28544:
DT 19-MAR-1993 (first entry)
DE CRI-4 (35E) analogue.
KW short consensus repeat; regulator of complement activation;
OS C3b binding; C4b binding; human complement type 1 receptor.
OS Homo sapiens.
FH Key
FT region Location/Qualifiers
FT /label= SCR-1
FT region 61..122
FT /label= SCR-2
FT region 451..510
FT /label= SCR-8
FT region 511..543
FT /label= SCR-9
FT /note= "TRUNCATED"
FT misc-difference 35
FT /note= "Gly substituted by Glu (SCR-8)"
PN EP-512733-A.
PD 11-NOV-1992.
PF 28-APR-1992: 303826.
PR 03-MAY-1991: US-695514.
PA (UNIV ) UNIV WASHINGTON.
PI Atkinson JP, Hourcade D, Krych M;
DR WPI: 92-375009/46.
PT Complement activity regulator protein analogues - useful for
PT treating auto-immune diseases, to suppress transplant rejection,
PT for diagnosis etc.
PS Claim 11: Fig 2 and R1810; 23pp; English.
CC The cDNA clone designated CRI-4 was described in J. Exp. Med. (1988)
CC 168:1255-1270. It encodes the first 8 and a half amino terminal
CC SCRs of CRI. The invention concerns analogues of "regulator of
CC complement activation" proteins or truncated, hybrid or recombinant
CC forms of them. CRI-4 is a preferred truncated form and a number of
CC specified substitution variants of it are claimed. Positions 35 and
CC 37 of SCR-1 and the corresponding positions in SCR-8 have been
CC identified as important in C4b binding. The specification does not
CC contain the CRI-4 sequence; the sequence given here was constructed
CC from the full-length CRI amino acid sequence having GENSEQ
CC accession number R1810 and descriptions in the disclosure.
CC Sequence 543 AA;

Query Match 100.0%; Score 186; DB 1; Length 543;
Best Local Similarity 100.0%; Pred. No. 1.5e-18;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVFELVGEPSICTSNDQVGIMSG 33
DB 154 CNPSSGGRKVFELVGEPSICTSNDQVGIMSG 186

RESULT 6
ID R28545 standard; peptide: 543 AA.
AC R28545:
DT 19-MAR-1993 (first entry)
DE CRI-4 (37Y) analogue.
KW short consensus repeat; regulator of complement activation;
OS C3b binding; C4b binding; human complement type 1 receptor.
OS Homo sapiens.
FH Key
FT region Location/Qualifiers
FT /label= SCR-1
FT region 61..122
FT /label= SCR-2
FT region 451..510
FT /label= SCR-8
FT region 511..543
FT /label= SCR-9
FT /note= "TRUNCATED"

```

```

FT misc-difference 37
FT /note= "Ser substituted by Tyr (SCR-8)"
PN EP-512733-A.
PD 11-NOV-1992.
PF 28-APR-1992: 303826.
PR 03-MAY-1991: US-695514.
PA (UNIV ) UNIV WASHINGTON.
PI Atkinson JP, Hourcade D, Krych M;
DR WPI: 92-375009/46.
PT Complement activity regulator protein analogues - useful for
PT treating auto-immune diseases, to suppress transplant rejection,
PT for diagnosis etc.
PS Claim 11: Fig 2 and R1810; 23pp; English.
CC The cDNA clone designated CRI-4 was described in J. Exp. Med. (1988)
CC 168:1255-1270. It encodes the first 8 and a half amino terminal
CC SCRs of CRI. The invention concerns analogues of "regulator of
CC complement activation" proteins or truncated, hybrid or recombinant
CC forms of them. CRI-4 is a preferred truncated form and a number of
CC specified substitution variants of it are claimed. Positions 35 and
CC 37 of SCR-1 and the corresponding positions in SCR-8 have been
CC identified as important in C4b binding. The specification does not
CC contain the CRI-4 sequence; the sequence given here was constructed
CC from the full-length CRI amino acid sequence having GENSEQ
CC accession number R1810 and descriptions in the disclosure.
CC Sequence 543 AA;

Query Match 100.0%; Score 186; DB 1; Length 543;
Best Local Similarity 100.0%; Pred. No. 1.5e-18;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVFELVGEPSICTSNDQVGIMSG 33
DB 154 CNPSSGGRKVFELVGEPSICTSNDQVGIMSG 186

RESULT 7
ID R28546 standard; peptide: 543 AA.
AC R28546:
DT 19-MAR-1993 (first entry)
DE CRI-4 (44T, 47D, 49L) analogue.
KW short consensus repeat; regulator of complement activation;
OS C3b binding; C4b binding; human complement type 1 receptor.
OS Homo sapiens.
FH Key
FT region Location/Qualifiers
FT /label= SCR-1
FT region 61..122
FT /label= SCR-2
FT region 451..510
FT /label= SCR-8
FT region 511..543
FT /label= SCR-9
FT /note= "TRUNCATED"
FT misc-difference 44
FT /note= "Ile substituted by Thr (SCR-8)"
FT misc-difference 47
FT /note= "Lys substituted by Asp (SCR-8)"
FT misc-difference 49
FT /note= "Ser substituted by Leu (SCR-8)"
FT misc-difference 49
FT /note= "Ser substituted by Tyr (SCR-8)"
PN EP-512733-A.
PD 11-NOV-1992.
PF 28-APR-1992: 303826.
PR 03-MAY-1991: US-695514.
PA (UNIV ) UNIV WASHINGTON.
PI Atkinson JP, Hourcade D, Krych M;
DR WPI: 92-375009/46.
PT Complement activity regulator protein analogues - useful for
PT treating auto-immune diseases, to suppress transplant rejection,
PT for diagnosis etc.
PS Claim 11: Fig 2 and R1810; 23pp; English.
CC The cDNA clone designated CRI-4 was described in J. Exp. Med. (1988)

```



PA (BRIG-) BRIGHAM AND WOMEN'S HOSPITAL.  
 PI Fearon DT, Klickstein LB, Wong WM, Carson GR, Hoh M, Concino MF,  
 PI Makrides SC, Marsh HC;  
 DR WPI: 91-132854/18.  
 DR N-PSDB: Q11642.  
 PT Human complement receptor type 1 gene, encoded proteins and  
 fragments - for treatment of immune disorders, myocardial infarct,  
 damage due to inflammation and in treatment of thrombosis  
 PS Claim 41; Fig 1; 234pp; English.  
 CC The invention also covers fragments of this protein which have the  
 ability to bind C3b and/or C4b, have cofactor I activity or can  
 inhibit C3 or C5 convertase activity. The full-length protein, or  
 CC its specified fragments are used to treat patients with immune  
 disorders or a disorder caused by inappropriate complement  
 CC activity. The protein is also used to treat thrombotic conditions  
 CC in humans and animals. See also Q11643.  
 SQ Sequence 2039 AA.

Query Match 100.0%; Score 186; DB 1; Length 2039;  
 Best Local Similarity 100.0%; Pred. No. 6, 4e-18;  
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVFELVGEPSICTSNDQVGITWSG 33  
 |||||  
 DB 195 CNPSSGGRKVFELVGEPSICTSNDQVGITWSG 227

RESULT 3  
 ID R11982 standard; Protein; 1537 AA.  
 AC R11982;  
 DT 25-JUN-1991 (first entry)  
 DE Partial human complement type 1 receptor.  
 KM complement system; C3b/C4b receptor; CRI; allergic reaction;  
 KW Immune response; long homologous repeat; LHR.  
 OS Homo sapiens.  
 FH Key  
 FT region Location/Qualifiers  
 FT 1..438  
 FT /label= LHR-B  
 FT 439..891  
 FT /label= LHR-C  
 FT 892..1341  
 FT /label= LHR-D  
 FT 1495..1498  
 FT /note= "positively-charged; preceded by hydrophobic  
 sequence"  
 FT 1521..1526  
 FT /note= "has 67 per cent homology to site of protein  
 kinase C phosphorylation in the EGF  
 receptor"

PN WO9105047-A.  
 PD 18-APR-1991.  
 PF 25-SEP-1990; 005454.  
 PR 26-SEP-1989; US-412745.  
 PR 26-SEP-1990; US-912349.  
 PA (TCEL-) T CELLS SCI INC.  
 PA (UJCO ) JOHNS HOPKINS UNIVERSITY.  
 PA (BRIG-) BRIGHAM AND WOMEN'S HOSPITAL.  
 PI Fearon DT, Klickstein LB, Wong WM, Carson GR, Hoh M, Concino MF,  
 PI Makrides SC, Marsh HC;  
 DR WPI: 91-132854/18.  
 DR N-PSDB: Q11643.  
 PT Human complement receptor type 1 gene, encoded proteins and  
 fragments - for treatment of immune disorders, myocardial infarct,  
 damage due to inflammation and in treatment of thrombosis  
 PS Disclosure; Fig 5; 234pp; English.  
 CC This sequence comprises three of the four tandem, direct, long  
 CC homologous repeats of the full-length F allele of CRI. LHR-A is  
 CC absent. Each LHR might represent a single C3b/C4b binding domain,  
 CC making the receptor multivalent. The LHR's are composed of 7 short  
 CC consensus repeats of 60-70 residues resembling the SCR's of other  
 CC C3/C4 binding proteins. The protein and fragments of it having C3b

CC and/or C4b binding activity can be used to treat immune disorders  
 CC or disorders involving inappropriate complement activity.  
 CC See also Q11642.  
 SQ Sequence 1537 AA.

Query Match 100.0%; Score 186; DB 1; Length 1537;  
 Best Local Similarity 100.0%; Pred. No. 4, 7e-18;  
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVFELVGEPSICTSNDQVGITWSG 33  
 |||||  
 DB 143 CNPSSGGRKVFELVGEPSICTSNDQVGITWSG 175

RESULT 4  
 ID R28543 standard; peptide; 543 AA.  
 AC R28543;  
 DT 19-MAR-1993 (first entry)  
 DE CRI-4 (35E, 37Y) analogue.  
 KW short consensus repeat; regulator of complement activation;  
 KW C3b binding; C4b binding; human complement type 1 receptor.  
 OS Homo sapiens.

FH Key  
 FT region Location/Qualifiers  
 FT 1..60  
 FT /label= SCR-1  
 FT 61..122  
 FT /label= SCR-2  
 FT 451..510  
 FT /label= SCR-8  
 FT 511..543  
 FT /label= SCR-9  
 FT /note= "TRUNCATED"  
 FT msc\_difference 35  
 FT /note= "Gly substituted by Glu (SCR-8)"  
 FT msc\_difference 37  
 FT /note= "Ser substituted by Tyr (SCR-8)"

PN EP-512733-A.  
 PD 11-NOV-1992.  
 PF 28-APR-1992; 303826.  
 PR 03-MAY-1991; US-695514.  
 PA (UNIV ) UNIV WASHINGTON.  
 PI Atkinson JP, Hourcade D, Kirsch M;  
 DR WPI: 92-375009/46.  
 PT Complement activity regulator protein analogues - useful for  
 treating auto-immune diseases, to suppress transplant rejection,  
 PT for diagnosis etc.  
 PS Claim 11; Fig 2 and R11810; 23pp; English.  
 CC The CDNA clone designated CRI-4 was described in J Exp Med. (1988)  
 CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
 CC SCRs of CRI. The invention concerns analogues of "regulator of  
 CC complement activation" proteins or truncated, hybrid or recombinant  
 CC forms of them. CRI-4 is a preferred truncated form and a number of  
 CC specified substitution variants of it are claimed. Positions 35 and  
 CC 37 of SCR-1 and the corresponding positions in SCR-8 have been  
 CC identified as important in C4b binding. The specification does not  
 CC contain the CRI-4 sequence; the sequence given here was constructed  
 CC from the full-length CRI amino acid sequence having GENSEQ  
 CC accession number R11810 and descriptions in the disclosure.  
 SQ Sequence 543 AA;

Query Match 100.0%; Score 186; DB 1; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 1, 5e-18;  
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVFELVGEPSICTSNDQVGITWSG 33  
 |||||  
 DB 154 CNPSSGGRKVFELVGEPSICTSNDQVGITWSG 186

RESULT 5

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 1999, 22:19:47 ; Search time 58.51 Seconds  
(without alignments)  
11.407 Million cell updates/sec

Title: US-09-142-043-1

Perfect score: 186  
Sequence: 1 CNPSSGGRKVFELVGEPSICTSNDQVGIWSG 33

Scoring table: BLOSUM62

Searched: 162890 seqs, 20225328 residues

Database: A.Geneseq\_34.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	186	100.0	2317	1 P92219	CRI protein New nu
2	186	100.0	2039	1 R1810	Human complement t
3	186	100.0	1537	1 R11982	Partial human comp
4	186	100.0	543	1 R28543	CRI-4 (35E, 37Y) a
5	186	100.0	543	1 R28544	CRI-4 (35E) analog
6	186	100.0	543	1 R28545	CRI-4 (37Y) analog
7	186	100.0	543	1 R28546	CRI-4 (44T, 47D, 4
8	186	100.0	543	1 R28547	CRI-4 (52S, 53S, 5
9	186	100.0	543	1 R28548	CRI-4 (57V, 59K) a
10	186	100.0	543	1 R28549	CRI-4 (64K, 65T) a
11	186	100.0	543	1 R28550	CRI-4 (64K) analog
12	186	100.0	543	1 R28551	CRI-4 (65T) analog
13	186	100.0	543	1 R28552	CRI-4 (78T, 79D) a
14	186	100.0	543	1 R28553	CRI-4 (85R, 87N) a
15	186	100.0	543	1 R28554	CRI-4 (92T, 94H) a
16	186	100.0	543	1 R28555	CRI-4 (92T) analog
17	186	100.0	543	1 R28556	CRI-4 (94H) analog
18	186	100.0	543	1 R28557	CRI-4 (99H, 103E)
19	186	100.0	543	1 R28558	CRI-4 (109N, 110A,
20	186	100.0	543	1 R28559	CRI-4 (114-117STP
21	186	100.0	543	1 R28560	CRI-4 (114S) analo
22	186	100.0	543	1 R28561	CRI-4 (115T) analo
23	186	100.0	543	1 R28562	CRI-4 (116K) analo
24	186	100.0	543	1 R28563	CRI-4 (117P) analo
25	186	100.0	543	1 R28564	CRI-4 (116K, 117P)
26	186	100.0	543	1 R28565	CRI-4 (121Q) analo
27	186	100.0	481	1 R29091	CRI-4 (amino acids
28	186	100.0	481	1 R29092	CRI-4 (amino acids
29	186	100.0	543	1 R28566	CRI-4 (318R, 319N)
30	186	100.0	543	1 R28567	CRI-4 (318-321 RNP
31	186	100.0	543	1 R28568	CRI-4 (347T, 349Y)
32	186	100.0	543	1 R28569	CRI-4 (369-376 STR
33	186	100.0	543	1 R28570	CRI-4 (266-274 KIK
34	186	100.0	543	1 R28571	CRI-4 (364-367 NNA
35	186	100.0	2039	1 R28572	CRI-4 (364-367 NNA
36	186	100.0	197	1 R36743	CRI-4 (364-367 NNA
37	186	100.0	76	1 R47153	Sequence of solubl
38	186	100.0	254	1 R47154	Sequence of solubl
39	186	100.0	254	1 R47155	Sequence of solubl
40	186	100.0	133	1 R47156	Sequence of solubl
41	186	100.0	211	1 W45909	SCR 1-3 of complen
42	186	100.0	209	1 W45912	SCR 1-3 of complen
43	186	100.0	198	1 W45897	SCR 1-3 of complen

## ALIGNMENTS

44	186	100.0	214	1 W45898	SCR 1-3 of complen
45	186	100.0	1930	1 W45899	Human complement r
ALIGNMENTS					
RESULT	1				
ID	P92219	standard; protein: 2317 AA.			
AC	P92219				
DE	22-FEB-1990 (first entry)				
OS	Complement; cofactor.				
OS	Homo sapiens (human).				
FT	Key	Location/Qualifiers			
FT	peptide	10..50			
FT		/label= signal_peptide			
PN	W08909220-A.				
PD	05-OCT-1989.				
PE	31-MAR-1989: U01358.				
PR	01-APR-1988: US-176532.				
PA	(TCEU) T Cell Sciences Inc; (UYJO) The Johns Hopkins University;				
PA	(BRIG*) The Brigham and Women's Hospital.				
PI	Fearon DT, Klickstein LB, Wong W, Carson G, Concino MF, Makrides SC;				
DR	WPI: 89-309498/42.				
DR	N-PSDB: N91477.				
PT	New nucleic acid sequences encoding new CRI protein - and its fragment,				
PT	for diagnosis and control of complement-related immune defects,				
PT	inflammation, myocardial infarct, etc				
PS	Claim 1; fig. 1; 191pp; English.				
CC	This is full-length CRI protein, and shortened forms are new, lacking				
CC	the transmembrane region. The proteins and fragments bind C3b and/or				
CC	C4b, have cofactor activity and inhibit C3 and C5 convertase activity.				
CC	In the sequence, x-untranslated region. This has 7 short consensus				
CC	repeats (CR) of 4 long direct homologous repeats (LHR) (see fig. 10 in				
CC	patent). Deletion mutants, eg LHR-A for binding of C4b, and LHR-C				
CC	for C3b binding, be made, and LHR-B and -C can detect CR2 sequences.				
CC	They are useful in diagnosing and treating immune disorders, and prevent				
CC	perfusion injury.				
SC	Sequence 2317 AA;				
Query Match					
Best local similarity 100.0%; Score 186; DB 1; Length 2317;					
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1 CNPSSGGRKVFELVGEPSICTSNDQVGIWSG 33				
DB	204 CNPSSGGRKVFELVGEPSICTSNDQVGIWSG 236				
RESULT	2				
ID	R1810	standard; protein: 2039 AA.			
AC	R1810				
DE	25-JUN-1991 (first entry)				
DE	Human complement type 1 receptor.				
KW	Complement system; C3b/C4b receptor; CRI; allergic reaction;				
KW	immune response; clone lambda T109.1.				
OS	Homo sapiens.				
FT	Key	Location/Qualifiers			
FT	peptide	1..41			
FT		/label= putative signal peptide			
FT	protein.	42..2039			
FT		/label= CRI			
PN	W09105047-A.				
PD	18-APR-1991.				
PE	25-SEP-1990; U05454.				
PR	26-SEP-1989; US-412745.				
PR	26-SEP-1990; US-912349.				
PA	(TCEU) T CELL SCI INC.				
PA	(UYJO) JOHNS HOPKINS UNIVERSITY.				

DE CRI-4 (64K, 65T) analogue.  
KW Short consensus repeat; regulator of complement activation;  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT region 1..60  
FT /label= SCR-1  
FT 61..122  
FT /label= SCR-2  
FT 451..510  
FT /label= SCR-8  
FT 511..543  
FT /label= SCR-9  
FT /note= "TRUNCATED"  
FT misc\_difference 64  
FT /note= "Arg substituted by Lys (SCR-9)"  
FT /note= "Asn substituted by Thr (SCR-9)"  
PN EP-512733-A.  
PD 11-NOV-1992.  
PF 28-APR-1992: 303826.  
PR 03-MAY-1991: US-695514.  
PA (UNIV ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Krych M;  
DR WPI: 92-375009/46.  
PT Complement activity regulator protein analogues - useful for  
PT treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11: Fig 2 and R1810: 23pp: English.  
CC The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCR-8 of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-2 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-9. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQ  
CC accession number R1810 and descriptions in the disclosure.  
CC Sequence 543 AA:  
SQ  
Query Match 100.0%; Score 186; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 1.5e-18;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 CNPSSGGRKVFELVGEPSICTSNDQVGIMSG 33  
DB 154 CNPSSGGRKVFELVGEPSICTSNDQVGIMSG 186  
RESULT 11  
R28550 ID R28550 standard: peptide; 543 AA.  
AC R28550;  
DT 19-MAR-1993 (first entry)  
DE CRI-4 (64K) analogue.  
KW Short consensus repeat; regulator of complement activation;  
KW C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT region 1..60  
FT /label= SCR-1  
FT 61..122  
FT /label= SCR-2  
FT 451..510  
FT /label= SCR-8  
FT 511..543  
FT /label= SCR-9  
FT /note= "TRUNCATED"  
FT misc\_difference 64

FT EP-512733-A. /note= "Arg substituted by Lys (SCR-9)"  
PN 11-NOV-1992.  
PD 28-APR-1992: 303826.  
PF 03-MAY-1991: US-695514.  
PA (UNIV ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Krych M;  
DR WPI: 92-375009/46.  
PT Complement activity regulator protein analogues - useful for  
PT treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11: Fig 2 and R1810: 23pp: English.  
CC The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCR-8 of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-2 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-9. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQ  
CC accession number R1810 and descriptions in the disclosure.  
CC Sequence 543 AA:  
SQ  
Query Match 100.0%; Score 186; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 1.5e-18;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 CNPSSGGRKVFELVGEPSICTSNDQVGIMSG 33  
DB 154 CNPSSGGRKVFELVGEPSICTSNDQVGIMSG 186  
RESULT 12  
R28551 ID R28551 standard: peptide; 543 AA.  
AC R28551;  
DT 19-MAR-1993 (first entry)  
DE CRI-4 (65T) analogue.  
KW Short consensus repeat; regulator of complement activation;  
KW C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT region 1..60  
FT /label= SCR-1  
FT 61..122  
FT /label= SCR-2  
FT 451..510  
FT /label= SCR-8  
FT 511..543  
FT /label= SCR-9  
FT /note= "TRUNCATED"  
FT misc\_difference 64  
FT /note= "Arg substituted by Lys (SCR-9)"  
FT /note= "Asn substituted by Thr (SCR-9)"  
PN EP-512733-A.  
PD 11-NOV-1992.  
PF 28-APR-1992: 303826.  
PR 03-MAY-1991: US-695514.  
PA (UNIV ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Krych M;  
DR WPI: 92-375009/46.  
PT Complement activity regulator protein analogues - useful for  
PT treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11: Fig 2 and R1810: 23pp: English.  
CC The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCR-8 of CRI. The invention concerns analogues of "regulator of

CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-2 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-9. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQ  
CC accession number R11810 and descriptions in the disclosure.  
SQ Sequence 543 AA;

Query Match 100.0%; Score 186; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 1.5e-18;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVEFLVGEPSICTSNDQVIGMSG 33  
DB 154 CNPSSGGRKVEFLVGEPSICTSNDQVIGMSG 186

## RESULT 13

R28552  
ID R28552 standard; peptide: 543 AA.  
AC R28552:  
DT 19-MAR-1993 (first entry)  
DE CRI-4 (78r, 79D) analogue.  
KW short consensus repeat; regulator of complement activation;  
KW C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.  
FH Key  
FT region Location/Qualifiers  
FT 1..60  
FT /label= SCR-1  
FT region 61..122  
FT /label= SCR-2  
FT region 451..510  
FT /label= SCR-8  
FT region 511..543  
FT /label= SCR-9  
FT /note= "TRUNCATED"  
FT misc\_difference 78  
FT /note= "Lys substituted by Thr (SCR-9)"  
FT misc\_difference 79  
FT /note= "Gly substituted by Asp (SCR-9)"  
PN EP-512733-A.  
PD 11-NOV-1992.  
PE 28-APR-1992; 303826.  
PR 03-MAR-1991; US-695514.  
PA (UNIW ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Krych M;  
WP1: 92-375009/46.  
DR WPI: 92-375009/46.  
PT Complement activator/regulator protein analogues - useful for  
PT treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11; Fig 2 and R11810; 23pp; English.  
CC The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCR-2 of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-2 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-9. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQ  
CC accession number R11810 and descriptions in the disclosure.  
SQ Sequence 543 AA;

Query Match 100.0%; Score 186; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 1.5e-18;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVEFLVGEPSICTSNDQVIGMSG 33  
DB 154 CNPSSGGRKVEFLVGEPSICTSNDQVIGMSG 186

## RESULT 14

R28553  
ID R28553 standard; peptide: 543 AA.  
AC R28553:  
DT 19-MAR-1993 (first entry)  
DE CRI-4 (85r, 87N) analogue.  
KW short consensus repeat; regulator of complement activation;  
KW C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.  
FH Key  
FT region Location/Qualifiers  
FT 1..60  
FT /label= SCR-1  
FT region 61..122  
FT /label= SCR-2  
FT region 451..510  
FT /label= SCR-8  
FT region 511..543  
FT /label= SCR-9  
FT /note= "TRUNCATED"  
FT misc\_difference 85  
FT /note= "Gln substituted by Arg (SCR-9)"  
FT misc\_difference 87  
FT /note= "Lys substituted by Asn (SCR-9)"  
PN EP-512733-A.  
PD 11-NOV-1992.  
PE 28-APR-1992; 303826.  
PR 03-MAR-1991; US-695514.  
PA (UNIW ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Krych M;  
WP1: 92-375009/46.  
DR WPI: 92-375009/46.  
PT Complement activator/regulator protein analogues - useful for  
PT treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11; Fig 2 and R11810; 23pp; English.  
CC The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCR-2 of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-2 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-9. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQ  
CC accession number R11810 and descriptions in the disclosure.  
SQ Sequence 543 AA;

Query Match 100.0%; Score 186; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 1.5e-18;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVEFLVGEPSICTSNDQVIGMSG 33  
DB 154 CNPSSGGRKVEFLVGEPSICTSNDQVIGMSG 186

## RESULT 15

R28554  
ID R28554 standard; peptide: 543 AA.  
AC R28554:  
DT 19-MAR-1993 (first entry)  
DE CRI-4 (92r, 94H) analogue.  
KW short consensus repeat; regulator of complement activation;  
KW C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.

```

FH Key Location/Qualifiers
FT region 1..60
FT /label= SCR-1
FT region 61..122
FT /label= SCR-2
FT region 451..510
FT /label= SCR-8
FT region 511..543
FT /label= SCR-9
FT /note= "TRUNCATED"
FT misc_difference 92
FT /note= "Lys substituted by Thr (SCR-9)"
FT misc_difference 94
FT /note= "Tyr substituted by His (SCR-9)"
PN EP-512733-A.
PD 11-NOV-1992.
PF 28-APR-1992; 303826.
PR 03-MAY-1991; US-695514.
PA (UNIV ) UNIV WASHINGTON.
PI Atkinson JP, Hourcade D, Krych M;
DR WPI; 92-375009/46.
PT Complement activity regulator protein analogues - useful for
PT treating auto-immune diseases, to suppress transplant rejection,
PT for diagnosis etc.
PS Claim 11, Fig 2 and R11810; 23pp; English.
CC The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)
CC 168:1255-1270. It encodes the first 8 and a half amino terminal
CC SCRs of CRI. The invention concerns analogues of "regulator of
CC complement activation" proteins or truncated, hybrid or recombinant
CC forms of them. CRI-4 is a preferred truncated form and a number of
CC specified substitution variants of it are claimed in which certain
CC positions in SCR-2 which have been identified as important for the
CC degree of C3b- and C4b-binding are substituted by amino acids from
CC the corresponding positions in SCR-9. The specification does not
CC contain the CRI-4 sequence; the sequence given here was constructed
CC from the full-length CRI amino acid sequence having GENESEQ
CC accession number R11810 and descriptions in the disclosure.
SQ Sequence 543 AA;

```

```

Query Match 100.0%; Score 186; DB 1; Length 543;
Best Local Similarity 100.0%; Pred.No.1.5e-18;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CNPSSGGRKVFELVGPISICTSNDQVGIMSG 33
   |||
Db 154 CNPSSGGRKVFELVGPISICTSNDQVGIMSG 186

```

Search completed: July 18, 1999, 00:42:52  
 Job time: 8585 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 Comugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 06:02:58 ; Search time 39.54 Seconds  
(without alignments)  
8.236 Million cell updates/sec

Title: US-09-142-043-1

Perfect score: 186  
Sequence: 1 CNPSSGGRKRVFLVGPSTICTSNDQVIGMSG 33

Scoring table: BLOSUM62

Searched: 106577 seqs, 9868381 residues

Database: Issued\_Patents\_AA:\*

1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/PCITUS9.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/Backfilest1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	186	100.0	197	2	US-08-356-361-27
2	186	100.0	76	2	US-08-356-361-28
3	186	100.0	254	2	US-08-356-361-29
4	186	100.0	254	2	US-08-356-361-30
5	186	100.0	133	2	US-08-356-361-31
6	186	100.0	197	2	US-08-769-967A-27
7	186	100.0	76	2	US-08-769-967A-28
8	186	100.0	254	2	US-08-769-967A-29
9	186	100.0	254	2	US-08-769-967A-30
10	186	100.0	133	2	US-08-769-967A-31
11	81.5	43.8	324	1	US-08-310-416A-14
12	81.5	43.8	324	1	US-08-888-171-14
13	81.5	43.8	323	2	US-08-435-149-2
14	81.5	43.8	577	2	US-08-435-149-3
15	60.5	32.5	254	1	US-08-310-416A-13
16	60.5	32.5	293	1	US-08-310-416A-16
17	60.5	32.5	169	1	US-08-310-416A-18
18	60.5	32.5	377	2	US-08-528-057-2
19	60.5	32.5	370	2	US-08-528-057-42
20	60.5	32.5	373	2	US-08-528-057-44
21	60.5	32.5	324	2	US-08-528-057-46
22	60.5	32.5	254	2	US-08-888-171-13
23	60.5	32.5	293	2	US-08-888-171-16
24	60.5	32.5	169	2	US-08-888-171-18
25	60.5	32.5	254	2	US-08-435-149-1
26	58	31.2	340	4	5256642-2
27	58	31.2	340	4	5472939-2
28	55	29.6	310	4	5256642-10
29	55	29.6	310	4	5472939-10
30	50.5	27.2	1196	1	US-08-144-121-4
31	50	26.9	50	4	5256642-5
32	50	26.9	250	4	5256642-6
33	50	26.9	260	4	5472939-5
34	50	26.9	250	4	5472939-6
35	49	26.3	10	4	5256642-18
36	49	26.3	10	4	5472939-18
37	48.5	26.1	610	1	US-08-365-470-3
38	47	25.3	387	1	US-08-713-828-3
39	47	25.3	387	2	US-08-919-627-3

40	45.5	24.5	60	1	US-08-210-266A-10	Sequence 10, Appl
41	45.5	24.5	60	1	US-08-688-675-10	Sequence 10, Appl
42	45	24.2	263	1	US-07-906-983-2	Sequence 2, Appl
43	45	24.2	385	2	US-08-340-539A-2	Sequence 2, Appl
44	45	24.2	385	2	US-08-461-592B-2	Sequence 2, Appl
45	45	24.2	372	2	US-08-513-278-2	Sequence 2, Appl

## ALIGNMENTS

```
RESULT 1
US-08-356-361-27
; Sequence 27, Application US/08356361
; Patent No. 5833989
; GENERAL INFORMATION:
; APPLICANT: Smith, Richard A.G.
; APPLICANT: Dodd, Ian
; APPLICANT: Freeman Mary A.
; APPLICANT: Mossakowska, Danuta E.I.
; TITLE OF INVENTION: No. 5833989el Compounds
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham-Corporate Intellectual Property
; STREET: P.O. Box 1539
; CITY: King of Prussia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/356,361
; FILING DATE: 03-Jul-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jervis, Herbert H.
; REGISTRATION NUMBER: 31,171
; REFERENCE/DOCKET NUMBER: P30423
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (610) 270-5019
; TELEFAX: (610) 270-5090
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 197 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: N-terminal
; US-08-356-361-27

Query Match 100.0% Score 186; DB 2; Length 197;
Best Local Similarity 100.0%; Pred. NO. 1.3e-16;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CNPSSGGRKRVFLVGPSTICTSNDQVIGMSG 33
DB 155 CNPSSGGRKRVFLVGPSTICTSNDQVIGMSG 187

RESULT 2
US-08-356-361-28
; Sequence 28, Application US/08356361
; Patent No. 5833989
; GENERAL INFORMATION:
; APPLICANT: Smith, Richard A.G.
; APPLICANT: Dodd, Ian
; APPLICANT: Freeman Mary A.
; APPLICANT: Mossakowska, Danuta E.I.
```

TITLE OF INVENTION: No. 5833989el Compounds  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: SmithKline Beecham-Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,361  
FILING DATE: 03-Jul-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jervis, Herbert H.  
REGISTRATION NUMBER: 31,171  
REFERENCE/DOCKET NUMBER: P30423  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5019  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 76 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: Internal  
US-08-356-361-28

Query Match 100.0%; Score 186; DB 2; Length 76;  
Best Local Similarity 100.0%; Pred. No. 4.4e-19;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CNPSSGGRKVFELVGEPSIYCTSNDDVGIMSG 33  
Db 34 CNPSSGGRKVFELVGEPSIYCTSNDDVGIMSG 66

RESULT 3  
US-08-356-361-29  
Sequence 29, Application US/08356361  
Patent No. 5833989  
GENERAL INFORMATION:  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman Mary A.  
APPLICANT: Mossakowska, Danuta E.I.  
TITLE OF INVENTION: No. 5833989el Compounds  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: SmithKline Beecham-Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,361  
FILING DATE: 03-Jul-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jervis, Herbert H.

REGISTRATION NUMBER: 31,171  
REFERENCE/DOCKET NUMBER: P30423  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5019  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-356-361-29

Query Match 100.0%; Score 186; DB 2; Length 254;  
Best Local Similarity 100.0%; Pred. No. 1.7e-18;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CNPSSGGRKVFELVGEPSIYCTSNDDVGIMSG 33  
Db 155 CNPSSGGRKVFELVGEPSIYCTSNDDVGIMSG 187

RESULT 4  
US-08-356-361-30  
Sequence 30, Application US/08356361  
Patent No. 5833989  
GENERAL INFORMATION:  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman Mary A.  
APPLICANT: Mossakowska, Danuta E.I.  
TITLE OF INVENTION: No. 5833989el Compounds  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: SmithKline Beecham-Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,361  
FILING DATE: 03-Jul-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jervis, Herbert H.  
REGISTRATION NUMBER: 31,171  
REFERENCE/DOCKET NUMBER: P30423  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5019  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-356-361-30

Query Match 100.0%; Score 186; DB 2; Length 254;  
Best Local Similarity 100.0%; Pred. No. 1.7e-18;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CNPSSGGRKVFELVGEPSIYCTSNDDVGIMSG 33



Db 155 CNPSSGGRKVFELVGEPSICTSNDQVGIWSG 187

## RESULT 5

US-08-356-361-31  
Sequence 31, Application US/08356361  
Patent No. 5833989  
GENERAL INFORMATION:  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman Mary A.  
APPLICANT: Mossakowska, Danuta E.I.  
TITLE OF INVENTION: No. 5833989el Compounds  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Smithline Beecham-Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,361  
FILING DATE: 03-Jul-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jarvis, Herbert H.  
REGISTRATION NUMBER: 31,171  
REFERENCE/DOCKET NUMBER: P30423  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5019  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 133 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-356-361-31

Query Match 100.0%; Score 186; DB 2; Length 133;  
Best Local Similarity 100.0%; Pred. No. 8.3e-19;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CNPSSGGRKVFELVGEPSICTSNDQVGIWSG 33  
Db 34 CNPSSGGRKVFELVGEPSICTSNDQVGIWSG 66

## RESULT 6

US-08-769-967A-27  
Sequence 27, Application US/08769967A  
Patent No. 5859223  
GENERAL INFORMATION:  
APPLICANT: Mossakowska, Danuta E.I.  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman, Anne Mary  
TITLE OF INVENTION: Soluble Crl Derivatives  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Smithline Beecham Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania

COUNTRY: USA  
ZIP: 19406

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,967A  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,569  
FILING DATE: 15-May-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: P30423C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5364  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 197 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-769-967A-27

Query Match 100.0%; Score 186; DB 2; Length 197;  
Best Local Similarity 100.0%; Pred. No. 1.3e-18;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CNPSSGGRKVFELVGEPSICTSNDQVGIWSG 33  
Db 155 CNPSSGGRKVFELVGEPSICTSNDQVGIWSG 187

## RESULT 7

US-08-769-967A-28  
Sequence 28, Application US/08769967A  
Patent No. 5859223  
GENERAL INFORMATION:  
APPLICANT: Mossakowska, Danuta E.I.  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman, Anne Mary  
TITLE OF INVENTION: Soluble Crl Derivatives  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Smithline Beecham Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,967A  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,569  
FILING DATE: 15-May-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954

REFERENCE/DOCKET NUMBER: P30423C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5364  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 76 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-769-967A-28

Query Match 100.0%; Score 186; DB 2; Length 76;  
Best Local Similarity 100.0%; Pred. No. 4,4e-19;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CNPSSGGRKVFELVGEPSICTSNDQVGIMSG 33  
DB 34 CNPSSGGRKVFELVGEPSICTSNDQVGIMSG 66

RESULT 8  
US-08-769-967A-29  
Sequence 29, Application US/08769967A  
Patent No. 5859223  
GENERAL INFORMATION:  
APPLICANT: Mossakowska, Danuta E.I.  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman, Anne Mary  
TITLE OF INVENTION: Soluble CRI Derivatives  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESS: Smithline Beecham Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,967A  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,569  
FILING DATE: 15-May-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: P30423C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5364  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-769-967A-29

Query Match 100.0%; Score 186; DB 2; Length 254;  
Best Local Similarity 100.0%; Pred. No. 1.7e-18;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CNPSSGGRKVFELVGEPSICTSNDQVGIMSG 33  
DB 155 CNPSSGGRKVFELVGEPSICTSNDQVGIMSG 187

RESULT 9  
US-08-769-967A-30  
Sequence 30, Application US/08769967A  
Patent No. 5859223  
GENERAL INFORMATION:  
APPLICANT: Mossakowska, Danuta E.I.  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman, Anne Mary  
TITLE OF INVENTION: Soluble CRI Derivatives  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESS: Smithline Beecham Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,967A  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,569  
FILING DATE: 15-May-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: P30423C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5364  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-769-967A-30

Query Match 100.0%; Score 186; DB 2; Length 254;  
Best Local Similarity 100.0%; Pred. No. 1.7e-18;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CNPSSGGRKVFELVGEPSICTSNDQVGIMSG 33  
DB 155 CNPSSGGRKVFELVGEPSICTSNDQVGIMSG 187

RESULT 10  
US-08-769-967A-31  
Sequence 31, Application US/08769967A  
Patent No. 5859223  
GENERAL INFORMATION:  
APPLICANT: Mossakowska, Danuta E.I.  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman, Anne Mary  
TITLE OF INVENTION: Soluble CRI Derivatives  
NUMBER OF SEQUENCES: 33

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: SmithKline Beecham Corporate Intellectual Property  
;; STREET: P.O. Box 1539  
;; CITY: King of Prussia  
;; STATE: Pennsylvania  
;; COUNTRY: USA  
;; ZIP: 19406  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/769,967A  
;; FILING DATE:  
;; CLASSIFICATION: 536  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/440,569  
;; FILING DATE: 15-May-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: King, William T.  
;; REGISTRATION NUMBER: 30,954  
;; REFERENCE/DOCKET NUMBER: P30423C2  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (610) 270-5364  
;; TELEFAX: (610) 270-5090  
;; INFORMATION FOR SEQ ID NO: 31:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 133 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; FRAGMENT TYPE: internal  
;; US-08-769-967A-31  
;;  
Query Match 100.0%; Score 186; DB 2; Length 133;  
Best Local Similarity 100.0%; Pred. No. 8.3e-19;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 CNPGSGGRKVFELVGPSTICTSNDQVGIMSG 33  
DB 34 CNPGSGGRKVFELVGPSTICTSNDQVGIMSG 66  
;;  
RESULT 11  
US-08-310-416A-14  
;; Sequence 14, Application US/08310416A  
;; Patent No. 5679546  
;; GENERAL INFORMATION:  
;; APPLICANT: Jone-Long Ko et al.  
;; TITLE OF INVENTION: CHIMERIC PROTEINS WHICH BLOCK  
;; NUMBER OF SEQUENCES: 19  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Fish & Richardson P.C.  
;; STREET: 225 Franklin Street  
;; CITY: Boston  
;; STATE: Massachusetts  
;; COUNTRY: U.S.A.  
;; ZIP: 02110-2804  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
;; COMPUTER: IBM PS/2 Model 502 or 55SX  
;; OPERATING SYSTEM: MS-DOS (Version 5.0)  
;; SOFTWARE: WordPerfect (Version 5.1)  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/310,416A  
;; FILING DATE: 22-SEP-1994  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Paul T. Clark  
;; REGISTRATION NUMBER: 30,162

;; REFERENCE/DOCKET NUMBER: 06180/005001  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617) 542-5070  
;; TELEFAX: (617) 542-8906  
;; TELEX: 200154  
;; INFORMATION FOR SEQ ID NO: 14:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 324 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: not relevant  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-310-416A-14  
;;  
Query Match 43.8%; Score 81.5; DB 1; Length 324;  
Best Local Similarity 51.5%; Pred. No. 0.00056;  
Matches 17; Conservative 4; Mismatches 5; Indels 7; Gaps 2;  
OY 1 CNPGSGGRKVFELVGPSTICTSNDQVGIMSG 33  
DB 219 CNKG-----FTMIGSHSITCTVNDG-GEWSG 244  
;;  
RESULT 12  
US-08-888-171-14  
;; Sequence 14, Application US/08888171  
;; Patent No. 5651528  
;; GENERAL INFORMATION:  
;; APPLICANT: Jone-Long, Ko  
;; APPLICANT: Higgins, Paul J.  
;; APPLICANT: Yeh, C. Grace  
;; TITLE OF INVENTION: METHODS OF INHIBITING COMPLEMENT  
;; NUMBER OF SEQUENCES: 19  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Fish & Richardson, P.C.  
;; STREET: 225 Franklin Street  
;; CITY: Boston  
;; STATE: MA  
;; COUNTRY: US  
;; ZIP: 02110-2804  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; COMPUTER: IBM Compatible  
;; OPERATING SYSTEM: Windows95  
;; SOFTWARE: FastSeq for Windows Version 2.0  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/888,171  
;; FILING DATE: 03-JUL-1997  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/310,416  
;; FILING DATE: 22-SEP-1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Freeman, John W.  
;; REGISTRATION NUMBER: 29,066  
;; REFERENCE/DOCKET NUMBER: 06180/005002  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 617/542-507  
;; TELEFAX: 617/542-890  
;; TELEX: 200154  
;;  
;; INFORMATION FOR SEQ ID NO: 14:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 324 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-888-171-14  
;;  
Query Match 43.8%; Score 81.5; DB 2; Length 324;  
Best Local Similarity 51.5%; Pred. No. 0.00056;  
Matches 17; Conservative 4; Mismatches 5; Indels 7; Gaps 2;

QY 1 CNPMSGKRVFELVGPSTYCTSDNDVGIWSG 33  
||| :||| :||| :||| :||| :|||  
Db 219 CNKG-----FTWIGHSTICTVNNDE-GEWSG 244

RESULT 13  
US-08-435-149-2  
; Sequence 2, Application US/08435149  
; Patent No. 5866402  
; GENERAL INFORMATION:  
; APPLICANT: INNIS, MICHAEL A.  
; APPLICANT: ZAROR, ISABEL  
; APPLICANT: CREASEY, ABELA A.  
; TITLE OF INVENTION: CHIMERIC MCP AND DAF PROTEINS WITH CELL  
; TITLE OF INVENTION: SURFACE LOCALIZING DOMAIN  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CHIRON CORPORATION  
; STREET: INTELLECTUAL PROPERTY - R440, P.O. BOX 8097  
; CITY: EMERYVILLE  
; STATE: CALIFORNIA  
; COUNTRY: U.S.A.  
; ZIP: 94662-8097  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/435,149  
; FILING DATE: 05-MAY-1995  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SAVERIDE, PAUL B.  
; REGISTRATION NUMBER: 36,914  
; REFERENCE/DOCKET NUMBER: 0989.001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 601-2585  
; TELEFAX: (510) 655-3542  
; TELEX: N/A  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 323 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-435-149-2

Query Match 43.8%; Score 81.5; DB 2; Length 323;  
Best Local Similarity 51.5%; Pred. No. 0.00056;  
Matches 17; Conservative 4; Mismatches 5; Indels 7; Gaps 2;

QY 1 CNPMSGKRVFELVGPSTYCTSDNDVGIWSG 33  
||| :||| :||| :||| :||| :|||  
Db 219 CNKG-----FTWIGHSTICTVNNDE-GEWSG 244

RESULT 14  
US-08-435-149-3  
; Sequence 3, Application US/08435149  
; Patent No. 5866402  
; GENERAL INFORMATION:  
; APPLICANT: INNIS, MICHAEL A.  
; APPLICANT: ZAROR, ISABEL  
; APPLICANT: CREASEY, ABELA A.  
; TITLE OF INVENTION: CHIMERIC MCP AND DAF PROTEINS WITH CELL  
; TITLE OF INVENTION: SURFACE LOCALIZING DOMAIN  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CHIRON CORPORATION

STREET: INTELLECTUAL PROPERTY - R440, P.O. BOX 8097  
CITY: EMERYVILLE  
STATE: CALIFORNIA  
COUNTRY: U.S.A.  
ZIP: 94662-8097  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/435,149  
; FILING DATE: 05-MAY-1995  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SAVERIDE, PAUL B.  
; REGISTRATION NUMBER: 36,914  
; REFERENCE/DOCKET NUMBER: 0989.001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 601-2585  
; TELEFAX: (510) 655-3542  
; TELEX: N/A  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 577 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-435-149-3

Query Match 43.8%; Score 81.5; DB 2; Length 577;  
Best Local Similarity 51.5%; Pred. No. 0.0011;  
Matches 17; Conservative 4; Mismatches 5; Indels 7; Gaps 2;

QY 1 CNPMSGKRVFELVGPSTYCTSDNDVGIWSG 33  
||| :||| :||| :||| :||| :|||  
Db 473 CNKG-----FTWIGHSTICTVNNDE-GEWSG 498

RESULT 15  
US-08-310-416A-13  
; Sequence 13, Application US/08310416A  
; Patent No. 5679546  
; GENERAL INFORMATION:  
; APPLICANT: Jone-Long Ko et al.  
; TITLE OF INVENTION: CHIMERIC PROTEINS WHICH BLOCK  
; TITLE OF INVENTION: COMPLEMENT ACTIVATION  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; COMPUTER: IBM PS/2 Model 502 or 555X  
; OPERATING SYSTEM: MS-DOS (Version 5.0)  
; SOFTWARE: WordPerfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/310,416A  
; FILING DATE: 22-SEP-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Paul T. Clark  
; REGISTRATION NUMBER: 30,162  
; REFERENCE/DOCKET NUMBER: 06180/005001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906

TELEX: 200154  
 INFORMATION FOR SEQ ID NO: 13:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 254 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: not relevant  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-310-416A-13

Query Match 32.5%; Score 60.5; DB 1; Length 254;  
 Best Local Similarity 40.6%; Pred. No. 0.33;  
 Matches 13; Conservative 5; Mismatches 9; Indels 5; Gaps 2;

OY 1 CNPSSGGRKVFELVGEPSITCTSNDDQVGIWS 32  
 DB 157 CDP-APGPDPFSLIGESTIYCGDN---SWWS 183

Search completed: July 18, 1999, 06:07:55  
 Job time: 297 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 06:02:54 ; Search time 44.47 Seconds  
(without alignments)  
27.798 Million cell updates/sec

Title: US-09-142-043-1

Perfect score: 186

Sequence: 1 CNPGSGGRKRVFELVGEPSITCTSDNDQVGIMSG 33

Scoring table: BLOSUM62

Searched: 116738 seqs, 37460341 residues

Database : PIR\_58:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	186	100.0	2014	2	I36936	complement recepto
2	186	100.0	661	2	I36937	complement recepto
3	186	100.0	2489	2	I73012	complement C3b/C4b
4	186	100.0	2039	2	A28507	complement C3b/C4b
5	159	85.5	482	2	A34924	complement C3b/C4b
6	129	69.4	433	2	A30550	complement C3b/C4b
7	129	69.4	440	2	A43519	complement recepto
8	127	68.3	497	2	UC2054	complement recepta
9	116	62.4	676	2	A45900	complement C3b rec
10	98	52.7	1091	1	PL0009	complement C3d/Eps
11	83.5	44.9	340	2	I56234	decay-accelerating
12	81.5	43.8	381	1	B26359	decay-accelerating
13	81.5	43.8	440	2	A26359	decay-accelerating
14	78	41.9	1025	2	A43526	complement C3d/Eps
15	73	39.2	597	2	S53711	C4b alpha chain p
16	71	38.2	330	2	I55975	X/Y protein - mous
17	64	34.4	469	1	NBMSCA	C4b-binding protei
18	61	32.8	597	1	NBHUC4	C4b-binding protei
19	60.5	32.5	369	2	I57998	membrane cofactor
20	60.5	32.5	384	2	S01896	membrane cofactor
21	60.5	32.5	377	2	I54479	membrane cofactor
22	60.5	32.5	349	2	G02913	sperm CD46 - human
23	60	32.3	558	2	A57953	C4b protein alpha
24	57	30.6	579	2	A56740	sperm egg recognit
25	56.5	30.4	369	2	UC5138	membrane cofactor
26	56.5	30.4	362	2	UC5194	membrane cofactor
27	53	28.5	610	2	I46001	C4b-binding protei
28	52.5	28.2	596	2	JC2539	beta-N-acetylgluco
29	52	28.0	1234	1	NBMSH	complement factor
30	51	27.4	482	2	A55683	L-furunculase (EC
31	51	27.4	634	2	UC5092	E-selectin - pig
32	50.5	27.2	1786	1	MMHUB1	laminin beta-1 cha
33	50.5	27.2	1786	1	MMMSB1	laminin beta-1 cha
34	50.5	27.2	711	2	A64523	outer membrane pro
35	50	26.9	1967	2	S64604	hypothetical prote
36	50	26.9	363	2	B45900	complement C3d/Eps
37	50	26.9	87	2	S20912	regulatory protein
38	50	26.9	87	2	S29614	whib protein - Str
39	50	26.9	545	2	I48951	oncofetal antigen

40 50 26.9 611 2 F70325 conserved hypothet  
41 49.5 26.6 315 2 A71550 hypothetical prote  
42 49.5 26.6 156 2 A48886 8-oxo-7,8-dihydrog  
43 49 26.3 387 1 KIRBPG phosphorylase kina  
44 49 26.3 583 2 S56680 beta-fructofuranos  
45 49 26.3 372 2 JC5377 L-selectin - hamad

## ALIGNMENTS

RESULT 1  
I36936  
complement receptor 1 - chimpanzee (fragment)  
C:Species: Pan troglodytes (chimpanzee)  
C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 07-Aug-1998  
C:Accession: I36936  
R:Birmingham, D.J.; Shen, X.P.; Hourcade, D.; Nickells, M.W.; Atkinson, J.P.  
J. Immunol. 153, 691-700, 1994  
A>Title: Primary sequence of an alternatively spliced form of CRL. Candidate for the  
A:Reference number: I36935; MUID:94292799  
A:Accession: I36936  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Residues: 1-2014 <RES>  
A:Molecule type: mRNA  
A:Cross-references: GB:I24920; NID:9551564; PID:9557725  
C:Superfamily: complement factor H repeat homology <FH01>  
F:18-74/Domain: complement factor H repeat homology <FH01>  
F:19-136/Domain: complement factor H repeat homology <FH02>  
F:141-207/Domain: complement factor H repeat homology <FH03>  
F:333-391/Domain: complement factor H repeat homology <FH12>  
F:1041-1107/Domain: complement factor H repeat homology <FH26>  
F:1749-1815/Domain: complement factor H repeat homology <FH27>

Query Match 100.0%; Score 186; DB 2; Length 2014;  
Best Local Similarity 100.0%; Pred. No. 4.2e-17;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CNPGSGGRKRVFELVGEPSITCTSDNDQVGIMSG 33  
DB 170 CNPGSGGRKRVFELVGEPSITCTSDNDQVGIMSG 202

RESULT 2  
I36937  
complement receptor 1 - chimpanzee (fragment)  
C:Species: Pan troglodytes (chimpanzee)  
C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 07-Aug-1998  
C:Accession: I36937  
R:Birmingham, D.J.; Shen, X.P.; Hourcade, D.; Nickells, M.W.; Atkinson, J.P.  
J. Immunol. 153, 691-700, 1994  
A>Title: Primary sequence of an alternatively spliced form of CRL. Candidate for the  
A:Reference number: I36935; MUID:94292799  
A:Accession: I36937  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-661 <RES>  
A:Cross-references: GB:I24921; NID:9557726; PID:9557727  
C:Superfamily: complement factor H repeat homology <FH01>  
F:18-74/Domain: complement factor H repeat homology <FH01>  
F:19-136/Domain: complement factor H repeat homology <FH02>  
F:141-207/Domain: complement factor H repeat homology <FH03>  
F:213-268/Domain: complement factor H repeat homology <FH04>  
F:333-391/Domain: complement factor H repeat homology <FH22>  
F:396-462/Domain: complement factor H repeat homology <FH27>

Query Match 100.0%; Score 186; DB 2; Length 661;  
Best Local Similarity 100.0%; Pred. No. 1.3e-17;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CNPGSGGRKRVFELVGEPSITCTSDNDQVGIMSG 33

Db 170 CNPSSGGRKVFELVGEPSITCTSDNDQVGIWSG 202

## RESULT 3

173012 Complement C3b/C4b receptor (allotype S) precursor - human  
N:Alternate names: complement receptor type 1 (CRI); surface glycoprotein CD35  
C:Species: Homo sapiens (man)  
C:Date: 02-Jul-1996 #sequence revision 02-Jul-1996 #text\_change 24-Sep-1998  
C:Accession: I73012; A47602; S03291  
R:VIK, D.P.; Wong, W.W.  
J. Immunol. 151, 6214-6224, 1993  
A:Title: Structure of the gene for the F allele of complement receptor type 1 and sequen  
A:Reference number: I56203; MUID:94065175  
A:Accession: I73012  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2489 <RES>  
A:Cross-references: GB:J17418; NID:g306678; PID:g451303  
R:Wong, W.W.; Cahill, J.M.; Rosen, M.D.; Kennedy, C.A.; Bonaccio, E.T.; Morris, M.J.; W  
J. Exp. Med. 169, 847-863, 1989  
A:Title: Structure of the human CRI gene. Molecular basis of the structural and quantita  
A:Reference number: A47602  
A:Accession: A47602  
A:Molecule type: DNA  
A:Residues: 1-41 <WON>  
A:Cross-references: GB:X14893  
R:Hourcade, D.; Miesner, D.R.; Atkinson, J.P.; Holers, V.M.  
J. Exp. Med. 168, 1235-1270, 1988  
A:Title: Identification of an alternative polyadenylation site in the human C3b/C4b rece  
type 1.  
A:Reference number: S03291; MUID:89010527  
A:Accession: S03291  
A:Molecule type: mRNA  
A:Residues: 26-584 <HOU>  
A:Cross-references: EMBL:X14362; NID:g30197; PID:g736240  
A:Experimental source: clone CRI-4  
C:Genetics:  
A:Gene: GDB:CR1; CD35  
A:Cross-references: GDB:119800; OMIM:120620  
A:Map position: 1932-1932  
A:Intons: 41/1; 101/1; 134/2; 163/1; 296/1; 355/1; 389/2; 418/1; 491/1; 551/1; 584/2; 6  
; 1484/2; 1513/1; 1646/1; 1705/1; 1733/2; 1768/1; 1844/1; 1904/1; 1937/2; 1966/1; 2099/1  
C:Keywords: duplication; glycoprotein; transmembrane protein  
F:1-41/Domain: signal sequence #status predicted <SIG>  
F:42-584/Product: complement C3b/C4b receptor, secreted #status predicted <MAT>  
F:43-99/Domain: complement factor H repeat homology <FH01>  
F:104-161/Domain: complement factor H repeat homology <FH02>  
F:166-232/Domain: complement factor H repeat homology <FH03>  
F:238-293/Domain: complement factor H repeat homology <FH04>  
F:297-353/Domain: complement factor H repeat homology <FH05>  
F:358-416/Domain: complement factor H repeat homology <FH06>  
F:421-487/Domain: complement factor H repeat homology <FH07>  
F:554-611/Domain: complement factor H repeat homology <FH0A>  
F:616-682/Domain: complement factor H repeat homology <FH0B>  
F:808-866/Domain: complement factor H repeat homology <FH0C>  
F:1004-1061/Domain: complement factor H repeat homology <FH0D>  
F:1066-1133/Domain: complement factor H repeat homology <FH10>  
F:1138-1193/Domain: complement factor H repeat homology <FH11>  
F:1197-1253/Domain: complement factor H repeat homology <FH12>  
F:1258-1316/Domain: complement factor H repeat homology <FH13>  
F:1321-1387/Domain: complement factor H repeat homology <FH14>  
F:1393-1449/Domain: complement factor H repeat homology <FH15>  
F:1454-1511/Domain: complement factor H repeat homology <FH16>  
F:1516-1582/Domain: complement factor H repeat homology <FH17>  
F:1588-1643/Domain: complement factor H repeat homology <FH18>  
F:1647-1703/Domain: complement factor H repeat homology <FH19>  
F:1708-1766/Domain: complement factor H repeat homology <FH20>  
F:1771-1837/Domain: complement factor H repeat homology <FH21>  
F:1846-1904/Domain: complement factor H repeat homology <FH22>  
F:1907-1964/Domain: complement factor H repeat homology <FH23>

F:166-2035/Domain: complement factor H repeat homology <FHXC>  
F:2100-2156/Domain: complement factor H repeat homology <FH24>  
F:2161-2219/Domain: complement factor H repeat homology <FHXD>

Query Match 100.0%; Score 186; DB 2; Length 2489;  
Best Local Similarity 100.0%; Pred. No. 5, 3e-17;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 CNPSSGGRKVFELVGEPSITCTSDNDQVGIWSG 33  
Db 195 CNPSSGGRKVFELVGEPSITCTSDNDQVGIWSG 227

## RESULT 4

A28507 Complement C3b/C4b receptor precursor, membrane-bound (allotype F) - human  
N:Alternate names: complement receptor type 1 (CRI); surface glycoprotein CD35  
C:Species: Homo sapiens (man)  
C:Date: 19-Nov-1988 #sequence revision 06-Sep-1996 #text\_change 10-Sep-1997  
C:Accession: S03843; A28507; I56203; A24748; B24748; C24748  
R:Klickstein, L.B.; Barrow, T.J.; Miletic, V.; Rabson, L.D.; Smith, J.A.; Fearon, D.T  
J. Exp. Med. 168, 1699-1717, 1988  
A:Title: Identification of distinct C3b and C4b recognition sites in the human C3b/C4  
A:Reference number: S03843; MUID:89035992  
A:Accession: S03843  
A:Status: translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-2039 <KLJ>  
A:Cross-references: EMBL:Y00816; NID:g30185; PID:g30186  
R:Klickstein, L.B.; Wong, W.W.; Smith, J.A.; Weis, J.H.; Wilson, J.G.; Fearon, D.T.  
J. Exp. Med. 165, 1095-1112, 1987  
A:Title: Human C3b/C4b receptor (CRI). Demonstration of long homologous repeating dom  
A:Reference number: A28507; MUID:87168191  
A:Accession: A28507  
A:Molecule type: mRNA  
A:Residues: 503-771, 'FV', 774-2039 <KL2>  
R:VIK, D.P.; Wong, W.W.  
J. Immunol. 151, 6214-6224, 1993  
A:Title: Structure of the gene for the F allele of complement receptor type 1 and seq  
A:Reference number: I56203; MUID:94065175  
A:Accession: I56203  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-683, 'X', 685-1021, 'X', 1023-1614, 'V', 1616-1826, 'R', 1828-1849, 'D', 1851-187  
R:Wong, W.W.; Klickstein, L.B.; Smith, J.A.; Weis, J.H.; Fearon, D.T.  
Proc. Natl. Acad. Sci. U.S.A. 82, 7711-7715, 1985  
A:Title: Identification of a partial cDNA clone for the human receptor for complement  
A:Reference number: A94073; MUID:86067975  
A:Accession: A24748  
A:Molecule type: mRNA  
A:Residues: 311-333; 729-745; 831-845 <WON>  
C:Genetics:  
A:Gene: GDB:CR1; CD35  
A:Cross-references: GDB:119800; OMIM:120620  
A:Map position: 1932-1932  
A:Intons: 41/1; 101/1; 134/2; 163/1; 296/1; 355/1; 389/2; 418/1; 491/1; 551/1; 584/2  
; 1487/2; 1516/1; 1649/1; 1708/1; 1742/2; 1771/1; 1847/1; 1906/1; 1966/1; 1976/1; 200  
C:Keywords: duplication; glycoprotein; transmembrane protein  
F:1-41/Domain: signal sequence #status predicted <SIG>  
F:42-2039/Product: complement C3b/C4b receptor, membrane-bound #status predicted <MAT  
F:43-99/Domain: complement factor H repeat homology <FH01>  
F:104-161/Domain: complement factor H repeat homology <FH02>  
F:166-232/Domain: complement factor H repeat homology <FH03>  
F:238-293/Domain: complement factor H repeat homology <FH04>  
F:297-353/Domain: complement factor H repeat homology <FH05>  
F:358-416/Domain: complement factor H repeat homology <FH06>  
F:421-487/Domain: complement factor H repeat homology <FH07>  
F:493-549/Domain: complement factor H repeat homology <FH07>  
F:554-611/Domain: complement factor H repeat homology <FH08>  
F:616-682/Domain: complement factor H repeat homology <FH09>



```

Oy 1 CNPGSGGKRYEELVGEPSIYCTSNDDQVGISG 33
    |||:|||||||  |||||
Db 195 CNLGRGKKYFELVGEPSIYCTSKDDQVGISG 22

```

RESULT 8  
JC2054  
complement regulatory protein, 512 antigen precursor - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 27-Jun-1994 #sequence (revision 27-Jun-1994 #text\_change 22-Nov-1996  
C;Accession: JC2054; PC2027  
C;Skurudaa, G.; Seno, H.; Dohi, N.; Takizawa, H.; Nonaka, M.; Okada, N.; Okada, H.

Biochem. Biophys. Res. Commun. 198, 819-826, 1994  
A:Title: Molecular cloning of the rat complement regulatory protein, 512 antigen.  
A:Reference number: JC2054  
A:Accession: JC2054  
A:Molecule type: mRNA  
A:Residues: 1-497 <SAK>  
A:Accession: PC2027  
A:Molecule type: protein  
A:Residues: 39-51 <SA2>  
C:Comment: This protein plays a critical role in protection against complement mediated  
C:Superfamily: complement factor H repeat homology  
C:Keywords: glycoprotein, transmembrane protein  
F:1-36/Domain: signal sequence #status predicted <SIG>  
F:37-488/Product: complement regulatory protein, 512 antigen #status predicted <MAT>  
F:38-94/Domain: complement factor H repeat homology <FH1>  
F:99-156/Domain: complement factor H repeat homology <FH2>  
F:161-227/Domain: complement factor H repeat homology <FH3>  
F:233-288/Domain: complement factor H repeat homology <FH4>  
F:294-332/Domain: complement factor H repeat homology <FH5>  
F:357-413/Domain: complement factor H repeat homology <FH6>  
F:421-444/Domain: transmembrane #status predicted <TM>  
F:247,331,346,450,482,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 68.3%; Score 127; DB 2; Length 497;  
Best Local Similarity 66.7%; Pred. No. 1e-09;  
Matches 22; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Oy 1 CNPSSGGRKVELVGPSTICTSNDQVIGWS 33  
Db 190 CNTDARGKRLFLVGPSTICTSIDGVGWS 222

RESULT 9  
A:Accession: A45900  
C:Complement C3b receptor type 2 long form precursor - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 13-Mar-1997  
C:Accession: A45900; 148306  
R:Kurtz, C.B.; O'Toole, E.; Christensen, S.M.; Wels, J.H.  
J. Immunol. 144, 3581-3591, 1990  
A:Title: The murine complement receptor gene family. IV. Alternative splicing of Cr2 gene  
A:Reference number: A45900  
A:Accession: A45900  
A:Molecule type: preliminary  
A:Status: preliminary  
A:Residues: 1-676 <KUR>  
A:Cross-references: GB:M36470  
A:Experimental source: clone 31-1  
R:Kim, Y.U.; Kinoshita, T.; Molina, H.; Hourcade, D.; Seya, T.; Wagner, L.M.; Holers, V.  
J. Exp. Med. 181, 151-159, 1995  
A:Title: Mouse complement regulatory protein Cr1y/p65 uses the specific mechanisms of bc  
A:Reference number: 148306; MUID:95105691  
A:Accession: 148306  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 21-367 <RES>  
A:Cross-references: EMBL:U17128; NID:9595980; PID:9595982  
C:Genetics:  
A:Gene: Cr2  
A:Introns: 80/1; 113/2; 142/1; 275/1; 333/1  
C:Superfamily: complement factor H repeat homology  
F:22-78/Domain: complement factor H repeat homology <FH01>  
F:83-140/Domain: complement factor H repeat homology <FH02>  
F:145-211/Domain: complement factor H repeat homology <FH03>  
F:217-272/Domain: complement factor H repeat homology <FH04>  
F:276-331/Domain: complement factor H repeat homology <FH05>  
F:336-394/Domain: complement factor H repeat homology <FH06>  
F:399-458/Domain: complement factor H repeat homology <FH07>  
F:467-523/Domain: complement factor H repeat homology <FH08>  
F:531-587/Domain: complement factor H repeat homology <FH09>  
F:592-648/Domain: complement factor H repeat homology <FH10>

Query Match 62.4%; Score 116; DB 2; Length 676;  
Best Local Similarity 62.5%; Pred. No. 4.3e-08;  
Matches 20; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Oy 1 CNPSSGGRKVELVGPSTICTSNDQVIGWS 32  
Db 174 CHTGKREKFLDLVGPSTICTSKDQVIGWN 205

## RESULT 10

PL0009  
C:Complement C3d/Epstein-Barr virus receptor precursor - human  
N:Alternate names: complement receptor 2; CR2/CD21  
C:Species: Homo sapiens (hmn)  
C:Date: 30-Jun-1992 #sequence\_revision 07-Jul-1995 #text\_change 05-Sep-1997  
C:Accession: J10028; A39958; A22036; A24319; C24319; D24319; F24319;  
R:Wels, J.; Tothaker, L.E.; Smith, J.A.; Wels, J.H.; Fearon, D.T.  
J. Exp. Med. 167, 1047-1066, 1988  
A:Title: Structure of the human B lymphocyte receptor for C3d and the Epstein-Barr v1  
A:Reference number: J10028; MUID:88171282  
A:Accession: J10028  
A:Molecule type: mRNA  
A:Residues: 1-1091 <WE1>  
A:Note: nucleotides 1566-1625 are missing from Figure 1; therefore, residues 522-542  
R:Moore, M.D.; Cooper, N.R.; Tack, B.F.; Nemerow, G.R.  
Proc. Natl. Acad. Sci. U.S.A. 84, 9194-9198, 1987  
A:Title: Molecular cloning of the cDNA encoding the Epstein-Barr virus/C3d receptor (C3dR)  
A:Reference number: A39958; MUID:86097434  
A:Accession: A39958  
A:Molecule type: mRNA  
A:Residues: 1-456, 'G', 457-644, 'R', 646-669, 'R', 671-816, 'NCSAEVILKWLIERAF', 835-840, 'L',  
A:Cross-references: GB:J03565; NID:9181919; PID:9181920  
R:Fujisaku, A.; Harley, J.B.; Frank, M.B.; Gruner, B.A.; Frazier, B.; Holers, V.M.  
J. Biol. Chem. 264, 2118-2125, 1989  
A:Title: Genomic organization and polymorphisms of the human C3d/Epstein-Barr virus r  
A:Reference number: A32036; MUID:89123277  
A:Accession: A32036  
A:Molecule type: mRNA  
A:Residues: 1-456, 'G', 457-658, 718-1050, 'I', 1052-1060, 'E', 1062-1091 <FU0>  
A:Cross-references: GB:J04463  
R:Wels, J.; Fearon, D.T.; Klinkstein, L.B.; Wong, W.W.; Richards, S.A.; De Bruyn Ko  
Proc. Natl. Acad. Sci. U.S.A. 83, 5639-5643, 1986  
A:Title: Identification of a partial cDNA clone for the C3d/Epstein-Barr virus recept  
of complement.  
A:Reference number: A94114; MUID:86287311  
A:Accession: A24319  
A:Molecule type: protein  
A:Residues: 226-230, 'XIDQ', 257-267, 332-341, 583-591, 'Q', 593, 'D', 595-596, 728-735 <WE2>  
A:Experimental source: B-lymphoblastoid cell lines SB and Raj1  
C:Genetics:  
A:Gene: GDB:CR2  
A:Cross-references: GDB:119802; OMIM:120650  
A:Map position: 1q32-1q32  
C:Superfamily: complement C3d/Epstein-Barr virus receptor; complement factor H repeat  
C:Keywords: alternative splicing; duplication; glycoprotein; transmembrane protein  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-1091/Product: complement receptor 2 (16-repeat form) #status predicted <MAT>  
F:21-658, 718-1091/Product: complement receptor 2 (15-repeat form) #status predicted <  
F:91-146/Domain: complement factor H repeat homology <FH02>  
F:154-210/Domain: complement factor H repeat homology <FH03>  
F:215-271/Domain: complement factor H repeat homology <FH04>  
F:276-342/Domain: complement factor H repeat homology <FH05>  
F:351-406/Domain: complement factor H repeat homology <FH06>  
F:410-465/Domain: complement factor H repeat homology <FH07>  
F:470-521/Domain: complement factor H repeat homology <FH08>  
F:526-592/Domain: complement factor H repeat homology <FH09>  
F:601-656/Domain: complement factor H repeat homology <FH10>  
F:660-716/Domain: complement factor H repeat homology <FH11>  
F:720-772/Domain: complement factor H repeat homology <FH12>  
F:777-837/Domain: complement factor H repeat homology <FH13>  
F:846-901/Domain: complement factor H repeat homology <FH14>

F:909-965/Domain: complement factor H repeat homology <FH15>  
F:970-1026/Domain: complement factor H repeat homology <FH16>  
F:1034-1056/Domain: transmembrane #status predicted <TM>  
F:1057-1091/Domain: intracellular #status predicted <INT>  
F:121,127,294,372,622,698,858,881,919/Binding site: carbohydrate (asn) (covalent) #statu

Query Match 52.7%; Score 98; DB 1; Length 1091;  
Best Local Similarity 54.5%; Pred. No. 2e-05;  
Matches 18; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 CNPSSGGRKVFELVGPSTICTSNDQVGTSG 33  
DB 555 CNPSSGGRKVFELVGPSTICTSNDQVGTSG 587

RESULT 11

156234  
decay-accelerating factor - orangutan (fragment)  
C:Species: Pongo pygmaeus (orangutan)  
C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 07-Aug-1998  
C:Accession: 156234  
R:Nickalls, M.W.; Alvarez, J.I.; Lublin, D.M.; Atkinson, J.P.  
J: Immunol. 152, 676-685, 1994  
A:Title: Characterization of DAF-2, a high molecular weight form of decay-accelerating f  
A:Reference number: 156234  
A:Accession: 156234  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-340 <RES>  
A:Cross-references: GB:S67775; NID:q459389; PID:q459390  
C:Superfamily: decay-accelerating factor; complement factor H repeat homology  
F:1-53/Domain: complement factor H repeat homology (fragment) <FH01>  
F:122-179/Domain: complement factor H repeat homology <FH02>  
F:184-242/Domain: complement factor H repeat homology <FH03>

Query Match 44.9%; Score 83.5; DB 2; Length 340;  
Best Local Similarity 51.5%; Pred. No. 0.00055;  
Matches 17; Conservative 4; Mismatches 5; Indels 7; Gaps 2;

QY 1 CNPSSGGRKVFELVGPSTICTSNDQVGTSG 33  
DB 212 CNKNG-----FTMIGHSIYCTVNDDE-GEWSG 237

RESULT 12

B26359  
decay-accelerating factor (GPI-anchored splice form) precursor - human  
N:Alternate names: CD55; DAF splice form 2; decay-accelerating factor membrane-bound fo  
C:Species: Homo sapiens (man)  
C>Date: 05-Oct-1988 #sequence\_revision 16-Aug-1996 #text\_change 24-Apr-1998  
C:Accession: B26359; A27666; A39101; I52594; I52564  
R:Caras, I.W.; Davitz, M.A.; Rhee, L.; Weddell, G.; Martin Jr., D.W.; Nussenzweig, V.  
Nature 325, 545-549, 1987  
A:Title: Cloning of decay-accelerating factor suggests novel use of splicing to generate  
A:Reference number: A26359; MUID:87115845  
A:Accession: B26359  
A:Molecule type: mRNA  
A:Residues: 1-381 <CAR>  
A:Cross-references: GB:M30142; NID:q181464; PID:q181465  
R:Medof, M.E.; Lublin, D.M.; Holers, V.M.; Ayers, D.J.; Getty, R.R.; Leykam, J.F.; Atkin  
Proc. Natl. Acad. Sci. U.S.A. 84, 2007-2011, 1987  
A:Title: Cloning and characterization of cDNAs encoding the complete sequence of decay-a  
A:Reference number: A27666; MUID:87115602  
A:Accession: A27666  
A:Molecule type: mRNA  
A:Residues: 6-79, 'T', 81-84, 'M', 86-381 <MED>  
A:Cross-references: GB:M15799; NID:q181462; PID:q181463  
R:Moran, P.; Raab, H.; Kohr, W.J.; Garas, I.W.  
J. Biol. Chem. 266, 1250-1257, 1991  
A:Title: Glycophospholipid membrane anchor attachment. Molecular analysis of the cleavag  
A:Reference number: A39101; MUID:91093238

A:Accession: A39101  
A:Molecule type: protein  
A:Residues: 338-352 <MOD>  
R:Lublin, D.M.; Mallinson, G.; Poole, J.; Reid, M.E.; Thompson, E.S.; Ferdman, B.R.;  
Blood 84, 1276-1282, 1994  
A:Title: Molecular basis of reduced or absent expression of decay-accelerating factor  
A:Reference number: 152594  
A:Accession: 152594  
A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA  
A:Residues: 194-198, 'L', 200-209 <LUB>  
A:Cross-references: GB:S72858; NID:q639599; PID:q639600  
A:Experimental source: individual KW, Cromer blood group phenotype Dr(a-)  
A:Note: the single nucleotide difference in this allele, which changes Ser-199 to Leu  
on (see reference 152564), and thus reduced DAF expression  
R:Reid, M.E.; Mallinson, G.; Sim, R.B.; Poole, J.; Pausch, V.; Merry, A.H.; Liew, Y.W.  
Blood 78, 3291-3297, 1991  
A:Title: Biochemical studies on red blood cells from a patient with the Inab phenotype  
A:Reference number: 152564; MUID:92075980  
A:Accession: 152564  
A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA  
A:Residues: 190-193, 'QUCPVE' <RE2>  
A:Cross-references: GB:S70688; NID:q240301; PID:q240302  
A:Experimental source: individual KW, Cromer blood group phenotype Dr(a-) (described  
C:Comment: Cromer blood group system antigens reside on this protein.  
C:Genetics:

A:Gene: GDB:DAF  
A:Cross-references: GDB:119088; OMIM:125240  
A:Map position: 1q32-1q32  
C:Function:

A:Description: protects tissues from damage by regulating complement activation on ce  
C:Superfamily: decay-accelerating factor; complement factor H repeat homology  
F:1-34/Domain: signal sequence #status predicted <SIG>  
F:35-353/Product: decay-accelerating factor 2 #status predicted <MAT>  
F:36-94/Domain: complement factor H repeat homology <FH01>  
F:98-158/Domain: complement factor H repeat homology <FH02>  
F:163-220/Domain: complement factor H repeat homology <FH03>  
F:225-283/Domain: complement factor H repeat homology <FH04>  
F:354-381/Domain: carboxyl-terminal propeptide #status predicted <PRO>  
F:95/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:353/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Ser) (in mature f

Query Match 43.8%; Score 81.5; DB 1; Length 381;  
Best Local Similarity 51.5%; Pred. No. 0.0012;  
Matches 17; Conservative 4; Mismatches 5; Indels 7; Gaps 2;

QY 1 CNPSSGGRKVFELVGPSTICTSNDQVGTSG 33  
DB 253 CNKNG-----FTMIGHSIYCTVNDDE-GEWSG 278

RESULT 13

A26359  
decay-accelerating factor splice form 1 precursor - human  
N:Alternate names: decay-accelerating factor 5; decay-accelerating factor secreted fo  
C:Species: Homo sapiens (man)  
C>Date: 05-Oct-1988 #sequence\_revision 05-Oct-1988 #text\_change 29-Aug-1997  
C:Accession: A26359; A39702; S16187; S23138; A27258  
R:Caras, I.W.; Davitz, M.A.; Rhee, L.; Weddell, G.; Martin Jr., D.W.; Nussenzweig, V.  
Nature 325, 545-549, 1987  
A:Title: Cloning of decay-accelerating factor suggests novel use of splicing to gener  
A:Reference number: A26359; MUID:87115845  
A:Accession: A26359  
A:Molecule type: mRNA  
A:Residues: 1-440 <CAR>  
A:Cross-references: GB:M30142  
R:Edwulonu, U.K.; Ravi, L.; Medof, M.E.  
Proc. Natl. Acad. Sci. U.S.A. 88, 4675-4679, 1991  
A:Title: Characterization of the decay-accelerating factor gene promoter region.  
A:Reference number: A39702; MUID:91271256

A:Accession: A39702  
A:Molecule type: DNA  
A:Residues: 1-79,'T',81-104 <EMW>  
A:Cross-references: GB:M64356  
A:Note: the authors translated the codon AGT for residue 85 as Met  
R:Nakano, T.; Sugita, Y.; Ishikawa, Y.; Choi, N.H.; Tobe, T.; Tomita, M.  
Biochim. Biophys. Acta 1074, 326-330, 1991  
A:Title: Isolation of two forms of decay-accelerating factor (DAF) from human urine.  
A:Reference number: S16187; MUID:91291869  
A:Accession: S16187  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 35-47 <BIO>  
R:Nakano, Y.; Sumida, K.; Kikuta, N.; Miura, N.H.; Tobe, T.; Tomita, M.  
Biochim. Biophys. Acta 1116, 235-240, 1992  
A:Title: Complete determination of disulfide bonds localized within the short consensus  
A:Reference number: S23138  
A:Accession: S23138  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 35-41,65-68,79-81,93-103,128-134,143-145,155-159,162-168,188-192,203-204,211  
R:Sugita, Y.; Negro, T.; Matsuda, T.; Sakamoto, T.; Tomita, M.  
J. Biochem. 100, 143-150, 1986  
A:Title: Improved method for the isolation and preliminary characterization of human DAF  
A:Reference number: A27258; MUID:87008461  
A:Accession: A27258  
A:Molecule type: protein  
A:Residues: 35,'X',37,'G',39-51,'P',53-55,'X',57-58,'X',60-63 <SUG>  
A:Note: Gly-35 and Leu-38 were also found  
C:Genetics:  
A:Gene: GDB:DAF  
A:Cross-references: GDB:119088; OMIM:125240  
A:Map position: 1q32-1q32  
C:Superfamily: decay-accelerating factor: complement factor H repeat homology  
C:Keywords: alternative splicing; glycoprotein  
F:1-34/Domain: signal sequence #status predicted <SIG>  
F:35-44/Product: decay-accelerating factor 1 #status predicted <MAT>  
F:36-94/Domain: complement factor H repeat homology <FH01>  
F:98-158/Domain: complement factor H repeat homology <FH02>  
F:163-220/Domain: complement factor H repeat homology <FH03>  
F:225-283/Domain: complement factor H repeat homology <FH04>  
F:95/Binding site: carbonylate (Asn) (covalent) #status predicted

Query Match 43.8%; Score 81.5; DB 2; Length 440;  
Best Local Similarity 51.5%; Pred. No: 0.0013;  
Matches 17; Conservative 4; Mismatches 5; Indels 7; Gaps 2;

OY 1 CNPSSGGRKRVELVGPSTICTSDNDQVGIWSG 33  
Db 253 CNKG-----FTMIGHSIYCTVNDE-GEWSG 278  
||| |::||| |||||::| | |||  
|::|| |::||| |::||| |::|||

RESULT 14  
A43526  
complement C3d/Epstein-Barr virus receptor 2 precursor - mouse  
N:Alternate names: complement receptor type 2  
C:Species: Mus musculus (house mouse)  
C>Date: 28-Oct-1992 #sequence-revision 28-Oct-1992 #text-change 31-Oct-1997  
C:Accession: A43526; A43538; A43215; A43802; B32215  
R:Fingerboth, J.D.  
J. Immunol. 144, 3458-3467, 1990  
A:Title: Comparative structure and evolution of murine CR2. The homolog of the human C3d  
A:Reference number: A43526; MUID:90229735  
A:Accession: A43526  
A:Molecule type: mRNA  
A:Residues: 1-1025 <FIN>  
A:Cross-references: GB:M35684; EMBL:J04153; NID:g192687; PID:g192688  
R:Molina, H.; Kinoshita, T.; Inoue, K.; Carel, J.C.; Holters, V.M.  
J. Immunol. 145, 2974-2983, 1990  
A:Title: A molecular and immunochemical characterization of mouse CR2. Evidence for a st  
A:Reference number: A43538; MUID:91010789  
A:Accession: A43538

A:Molecule type: mRNA  
A:Residues: 12-305,'T',307-519,'A',521-1025 <MOL>  
A:Cross-references: GB:M61132; NID:g192692; PID:g192693  
R:Fingerboth, J.D.; Benedict, M.A.; Levy, D.N.; Strominger, J.L.  
Proc. Natl. Acad. Sci. U.S.A. 86, 242-246, 1989  
A:Title: Identification of murine complement receptor type 2.  
A:Reference number: A32215; MUID:89098890  
A:Accession: A32215  
A:Molecule type: mRNA  
A:Residues: 343-401,991-1025 <FI2>  
A:Cross-references: GB:J04153  
R:Kurtz, C.B.; Paul, M.S.; Aegeerter, M.; Weis, J.J.; Weis, J.H.  
J. Immunol. 143, 2058-2067, 1989  
A:Title: Murine complement receptor gene family. Identification and characterization  
A:Reference number: A45802  
A:Accession: A45802  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 'E',100-101,292-961,964-1025 <KUR>  
A:Cross-references: GB:M29281; NID:g192685; PID:g387131  
A:Note: the authors failed to translate GGA for residue 421 as Gly, and CCA for resid  
A:Note: the authors translated the codon CAC for residue 727 as Asn  
C:Superfamily: complement C3d/Epstein-Barr virus receptor: complement factor H repeat  
C:Keywords: alternative splicing; duplication; glycoprotein; receptor; Transmembrane  
F:1-11/Domain: signal sequence #status predicted <SIG>  
F:12-1025/Product: complement C3d/Epstein-Barr virus receptor 2 (15-repeat form) #sta  
F:14-973/Domain: extracellular #status predicted <EXT>  
F:14-73/Domain: complement factor H repeat homology <FH01>  
F:82-138/Domain: complement factor H repeat homology <FH02>  
F:146-202/Domain: complement factor H repeat homology <FH03>  
F:207-263/Domain: complement factor H repeat homology <FH04>  
F:268-334/Domain: complement factor H repeat homology <FH05>  
F:343-398/Domain: complement factor H repeat homology <FH06>  
F:402-458/Domain: complement factor H repeat homology <FH07>  
F:463-514/Domain: complement factor H repeat homology <FH08>  
F:519-585/Domain: complement factor H repeat homology <FH09>  
F:594-649/Domain: complement factor H repeat homology <FH10>  
F:654-704/Domain: complement factor H repeat homology <FH12>  
F:709-769/Domain: complement factor H repeat homology <FH13>  
F:778-833/Domain: complement factor H repeat homology <FH17>  
F:841-887/Domain: complement factor H repeat homology <FH15>  
F:902-958/Domain: complement factor H repeat homology <FH16>  
F:968-989/Domain: transmembrane #status predicted <TMN>  
F:990-1025/Domain: intracellular #status predicted <INT>

Query Match 41.9%; Score 78; DB 2; Length 1025;  
Best Local Similarity 42.4%; Pred. No: 0.0099;  
Matches 14; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

OY 1 CNPSSGGRKRVELVGPSTICTSDNDQVGIWSG 33  
Db 297 CDPSPKGVSTFLIGKTICTTSQKTGIWSG 329  
||| |::||| |::||| |::||| |::|||  
|::|| |::||| |::||| |::|||

RESULT 15  
553711  
C4b alpha chain precursor - rabbit  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C>Date: 15-Jul-1995 #sequence-revision 21-Jul-1995 #text-change 07-Aug-1998  
C:Accession: S53711  
R:de Fritos, P.G.; Dahlbaeck, B.  
Biochim. Biophys. Acta 1261, 285-289, 1995  
A:Title: cDNA structure of rabbit C4b-binding protein alpha-chain. Preserved sequence  
A:Reference number: S53711  
A:Accession: S53711  
A:Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-597 <DEP>  
A:Cross-references: EMBL:Z35490  
C:Superfamily: C4b-binding protein alpha chain: complement factor H repeat homology  
F:50-107/Domain: complement factor H repeat homology <FH1>  
F:112-169/Domain: complement factor H repeat homology <FH2>

F:174-234/Domain: complement factor H repeat homology <FH3>  
F:239-294/Domain: complement factor H repeat homology <FH4>  
F:484-538/Domain: complement factor H repeat homology <FH8>

Query Match 39.2%; Score 73; DB 2; Length 597;  
Best Local Similarity 46.9%; Pred. No. 0.027;  
Matches 15; Conservative 3; Mismatches 8; Indels 6; Gaps 1;  
QY 1 CNPSSGSRKVFELVGEPSIYCTSNDDVGIMS 32  
||| ||:| | | | | : ||:| |  
Db 203 CNPR-----FSLGLEASISCTVKNKTIVGWS 228

Search completed: July 18, 1999, 06:07:02  
Job time: 248 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 11:08:41 ; Search time 31.53 Seconds

(without alignments)  
28.092 Million cell updates/sec

Title: US-09-142-043-1

Perfect score: 186

Sequence: 1 CNPMSGGRVFEVLGEPISYCTSNDDQVIGMSG 33

Scoring table: BLOSUM62

Searched: 74019 seqs, 26840295 residues

Database: SwissProt\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	186	100.0	2039	1	CRI_HUMAN	P17927 homo sapien
2	98	52.7	1033	1	CR2_HUMAN	P20023 homo sapien
3	83.5	44.9	340	1	DAF_PONPY	P49457 pongo pygma
4	81.5	43.8	381	1	DAF_HUMAN	P08174 homo sapien
5	78	41.9	1025	1	CR2_MOUSE	P19070 mus musculus
6	67	36.0	390	1	DAF1_MOUSE	P06147 mus musculus
7	64	34.4	469	1	C4BP_MOUSE	P08607 mus musculus
8	61	32.8	597	1	C4BP_HUMAN	P04003 homo sapien
9	60.5	32.5	377	1	MCP_HUMAN	P15529 homo sapien
10	60	32.3	558	1	C4BP_RAT	P06351 rattus norv
11	59	31.7	407	1	DAF2_MOUSE	P06146 mus musculus
12	57	30.6	958	1	HIG_DROME	P09101 drosophila
13	53	28.5	610	1	C4BP_BOVIN	P02805 bos taurus
14	52.5	28.2	507	1	DAF_CAVO	P06401 cavia porce
15	52.5	28.2	596	1	HEXC_BOWMO	P49010 bombyx mori
16	52	28.0	1234	1	CEA_MOUSE	P06909 mus musculus
17	52	28.0	210	1	CRB3_BOVIN	P19141 bos taurus
18	51	27.4	634	1	IDUA_MOUSE	P48441 mus musculus
19	51	27.4	484	1	LEM2_PIG	P07942 sus scrofa
20	50.5	27.2	1786	1	LMB1_HUMAN	P02469 homo sapien
21	50.5	27.2	1786	1	LMB1_MOUSE	P02469 homo sapien
22	50	26.9	590	1	INVA_MAIZE	P49174 zea mays (m
23	50	26.9	1967	1	YCSO_YEAST	P53337 saccharomyc
24	49.5	26.6	156	1	80DB_HUMAN	P36639 homo sapien
25	49	26.3	583	1	INV3_DAVCA	P09653 daucus caro
26	49	26.3	386	1	KPB6_RABIT	P00518 oryctolagus
27	49	26.3	372	1	LEM1_MACMU	P095198 macaca mula
28	49	26.3	372	1	LEM1_PAPHA	P28768 papio hamad
29	49	26.3	134	1	SPAK_SALTY	P39443 salmonella
30	48.5	26.1	610	1	LEM4_HUMAN	P16381 homo sapien
31	48	25.8	548	1	AMT4_PSEST	P13507 pseudomonas
32	48	25.8	460	1	ENV_HV123	P12491 human immun
33	48	25.8	539	1	PFB_LYSEN	P05205 lysobacter
34	48	25.8	401	1	YOVI_CAEEL	P09551 caenorhabdi
35	47.5	25.5	271	1	S24D_ANOGA	P17004 anopheles g
36	47.5	25.5	959	1	VILI_DICDI	P36418 dictyostell
37	47	25.3	789	1	CAH_MOUSE	P54212 dunaliella
38	47	25.3	560	1	CO2_MOUSE	P21180 mus musculus
39	47	25.3	386	1	KPB6_HUMAN	P08166 homo sapien
40	47	25.3	149	1	RHIC_RHILY	P03315 rhizobium l
41	46	24.7	829	1	CAD3_HUMAN	P22223 homo sapien
42	46	24.7	211	1	CRB3_BOVIN	P02524 rattus norv
43	46	24.7	646	1	LEM3_BOVIN	P42201 bos taurus

## ALIGNMENTS

RESULT	1	STANDARD:	PRT:	2039 AA.
CRI_HUMAN				
ID	CRI_HUMAN			
AC	P17927			
DT	01-NOV-1990 (REL. 16, CREATED)			
DT	01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)			
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)			
DE	COMPLEMENT RECEPTOR TYPE 1 PRECURSOR (C3B/C4B RECEPTOR) (CD35 ANTIGEN).			
GN	CRI OR C3BR.			
OS	HOMO SAPIENS (HUMAN).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;			
OC	EUTHERIA; PRIMATES.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 89035892.			
RA	KLICKSTEIN L.B., BARTOW T.J., MILETIC V., RABSON L.D., SMITH J.A.,			
RA	FEARON D.T.;			
RL	J. EXP. MED. 168:1699-1717(1988).			
RN	[2]			
RP	SEQUENCE OF 503-2039 FROM N.A.			
RX	MEDLINE: 87168191.			
RA	KLICKSTEIN L.B., WONG W.W., SMITH J.A., WEIS J.H., WILSON J.G.,			
RA	FEARON D.T.;			
RL	J. EXP. MED. 165:1095-1112(1987).			
RN	[3]			
RP	SEQUENCE OF 761-783; 831-845 AND 1179-1195 FROM N.A.			
RX	MEDLINE: 86067975.			
RA	WONG W.W., KLICKSTEIN L.B., SMITH J.A., WEIS J.H., FEARON D.T.;			
RA	PROC. NATL. ACAD. SCI. U.S.A. 82:7711-7715(1985).			
CC	-1- FUNCTION: CRI, PRESENT ON ERYTHROCYTES, LEUCOCYTES, GLOMERULAR			
CC	PODCYTES, AND SPLENIC FOLLICULAR DENDRITIC CELLS. MEDIATES THE			
CC	BINDING BY THESE CELLS OF PARTICLES AND IMMUNE COMPLEXES THAT HAVE			
CC	ACTIVATED COMPLEMENT.			
CC	-1- SUBUNIT: MONOMER.			
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.			
CC	-1- POLYPEPTIDE: CRI IS RESPONSIBLE FOR THE KNOPS BLOOD GROUP SYSTEM.			
CC	-1- SEVEN SHORT CONSENSUS REPEATS (SCR) CONSTITUTE A LONG HOMOLOGOUS			
CC	REPEAT (LHR). THE N-TERMINAL TWO SCRS OF LHR-A CONTAINED A SITE			
CC	DETERMINING C4 SPECIFICITY, AND THE N-TERMINAL TWO SCRS OF LHR-B			
CC	AND -C EACH HAD A SITE DETERMINING C3 SPECIFICITY.			
CC	-1- SIMILARITY: CONTAINS 30 SUSHI (SCR) REPEATS.			
DR	EMBL: M11569; G180995; -			
DR	EMBL: M11617; G180996; -			
DR	EMBL: M11618; G180997; -			
DR	EMBL: Y00816; G30186; -			
DR	EMBL: X05309; G809019; -			
DR	PIR: A28507; A28507.			
DR	PIR: A24748; A24748.			
DR	PIR: B24748; B24748.			
DR	PIR: C24748; C24748.			
DR	PIR: S03843; S03843.			
DR	HSSP: P08603; 1HFI.			
DR	MIM: 120620; -			
KW	COMPLEMENT PATHWAY; GLYCOPROTEIN; TRANSMEMBRANE; REPEAT; SIGNAL;			
KW	RECEPTOR; SUSHI; BLOOD GROUP ANTIGEN.			
FT	SIGNAL	1	41	
FT	CHAIN	1	2039	
FT	DOMAIN	42	1971	COMPLEMENT RECEPTOR TYPE 1.
FT	DOMAIN	1972	1996	EXTRACELLULAR (POTENTIAL).
FT	DOMAIN	1997	2039	POTENTIAL.
FT	MOD_RES	42	42	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	42	488	PYRROLIDONE CARBOXYLIC ACID (POTENTIAL).
FT	DOMAIN	42	488	7 X SUSHI (SCR) REPEATS, LONG HOMOLOGOUS
FT	REPEAT	42	100	REPEAT A.
FT	REPEAT	42	100	SUSHI A.

44 46 24.7 769 1 LEM3-SHEEP P98109 ovis aries  
45 46 24.7 243 1 NOEL\_RHISN P55473 rhizobium s

```

FT REPEAT 103 162 SUSHI A2.
FT REPEAT 165 233 SUSHI A3.
FT REPEAT 237 294 SUSHI A4.
FT REPEAT 296 354 SUSHI A5.
FT REPEAT 357 417 SUSHI A6.
FT REPEAT 420 488 SUSHI A7.
FT DOMAIN 492 938 7 X SUSHI
FT REPEAT 492 938 REPEAT B.
FT REPEAT 492 550 SUSHI B1.
FT REPEAT 553 612 SUSHI B2.
FT REPEAT 615 683 SUSHI B3.
FT REPEAT 687 744 SUSHI B4.
FT REPEAT 746 804 SUSHI B5.
FT REPEAT 807 867 SUSHI B6.
FT REPEAT 870 938 SUSHI B7.
FT DOMAIN 942 1388 7 X SUSHI (SCR) REPEATS, LONG HOMOLOGOUS
FT REPEAT 942 1000 REPEAT C.
FT REPEAT 1003 1062 SUSHI C1.
FT REPEAT 1065 1133 SUSHI C2.
FT REPEAT 1137 1194 SUSHI C3.
FT REPEAT 1196 1254 SUSHI C4.
FT REPEAT 1257 1317 SUSHI C5.
FT REPEAT 1320 1388 SUSHI C6.
FT DOMAIN 1395 1846 7 X SUSHI (SCR) REPEATS, LONG HOMOLOGOUS
FT REPEAT 1395 1453 REPEAT D.
FT REPEAT 1456 1515 SUSHI D1.
FT REPEAT 1518 1586 SUSHI D2.
FT REPEAT 1590 1647 SUSHI D3.
FT REPEAT 1649 1707 SUSHI D4.
FT REPEAT 1710 1770 SUSHI D5.
FT REPEAT 1773 1841 SUSHI D6.
FT DOMAIN 1847 1966 2 X SUSHI (SCR) REPEATS.
FT REPEAT 1847 1905 SUSHI D7.
FT REPEAT 1908 1966 SUSHI 1.
FT DISULFID 43 86 BY SIMILARITY.
FT DISULFID 73 99 BY SIMILARITY.
FT DISULFID 104 145 BY SIMILARITY.
FT DISULFID 131 161 BY SIMILARITY.
FT DISULFID 166 215 BY SIMILARITY.
FT DISULFID 195 232 BY SIMILARITY.
FT DISULFID 238 280 BY SIMILARITY.
FT DISULFID 266 293 BY SIMILARITY.
FT DISULFID 297 340 BY SIMILARITY.
FT DISULFID 326 353 BY SIMILARITY.
FT DISULFID 358 400 BY SIMILARITY.
FT DISULFID 386 416 BY SIMILARITY.
FT DISULFID 421 470 BY SIMILARITY.
FT DISULFID 450 487 BY SIMILARITY.
FT DISULFID 493 536 BY SIMILARITY.
FT DISULFID 523 549 BY SIMILARITY.
FT DISULFID 554 595 BY SIMILARITY.
FT DISULFID 581 611 BY SIMILARITY.
FT DISULFID 616 665 BY SIMILARITY.
FT DISULFID 646 682 BY SIMILARITY.
FT DISULFID 688 730 BY SIMILARITY.
FT DISULFID 716 743 BY SIMILARITY.
FT DISULFID 747 790 BY SIMILARITY.
FT DISULFID 776 803 BY SIMILARITY.
FT DISULFID 808 850 BY SIMILARITY.
FT DISULFID 836 866 BY SIMILARITY.
FT DISULFID 871 920 BY SIMILARITY.
FT DISULFID 900 937 BY SIMILARITY.
FT DISULFID 943 986 BY SIMILARITY.
FT DISULFID 973 999 BY SIMILARITY.
FT DISULFID 1004 1045 BY SIMILARITY.
FT DISULFID 1031 1061 BY SIMILARITY.
FT DISULFID 1066 1115 BY SIMILARITY.
FT DISULFID 1095 1132 BY SIMILARITY.
FT DISULFID 1138 1180 BY SIMILARITY.
FT DISULFID 1166 1193 BY SIMILARITY.
FT DISULFID 1197 1240 BY SIMILARITY.

```

```

FT DISULFID 1226 1253 BY SIMILARITY.
FT DISULFID 1258 1300 BY SIMILARITY.
FT DISULFID 1286 1316 BY SIMILARITY.
FT DISULFID 1321 1370 BY SIMILARITY.
FT DISULFID 1350 1387 BY SIMILARITY.
FT DISULFID 1396 1439 BY SIMILARITY.
FT DISULFID 1426 1452 BY SIMILARITY.
FT DISULFID 1457 1498 BY SIMILARITY.
FT DISULFID 1484 1514 BY SIMILARITY.
FT DISULFID 1519 1568 BY SIMILARITY.
FT DISULFID 1548 1585 BY SIMILARITY.
FT DISULFID 1591 1633 BY SIMILARITY.
FT DISULFID 1619 1646 BY SIMILARITY.
FT DISULFID 1650 1693 BY SIMILARITY.
FT DISULFID 1679 1706 BY SIMILARITY.
FT DISULFID 1711 1753 BY SIMILARITY.
FT DISULFID 1739 1769 BY SIMILARITY.
FT DISULFID 1774 1823 BY SIMILARITY.
FT DISULFID 1803 1840 BY SIMILARITY.
FT DISULFID 1848 1891 BY SIMILARITY.
FT DISULFID 1877 1904 BY SIMILARITY.
FT DISULFID 1909 1952 BY SIMILARITY.
FT DISULFID 1938 1965 BY SIMILARITY.
FT CARBOHYD 56 56 POTENTIAL.
FT CARBOHYD 252 252 POTENTIAL.
FT CARBOHYD 410 410 POTENTIAL.
FT CARBOHYD 447 447 POTENTIAL.
FT CARBOHYD 509 509 POTENTIAL.
FT CARBOHYD 578 578 POTENTIAL.
FT CARBOHYD 702 702 POTENTIAL.
FT CARBOHYD 860 860 POTENTIAL.
FT CARBOHYD 897 897 POTENTIAL.
FT CARBOHYD 959 959 POTENTIAL.
FT CARBOHYD 1028 1028 POTENTIAL.
FT CARBOHYD 1152 1152 POTENTIAL.
FT CARBOHYD 1310 1310 POTENTIAL.
FT CARBOHYD 1481 1481 POTENTIAL.
FT CARBOHYD 1504 1504 POTENTIAL.
FT CARBOHYD 1534 1534 POTENTIAL.
FT CARBOHYD 1540 1540 POTENTIAL.
FT CARBOHYD 1605 1605 POTENTIAL.
FT CARBOHYD 1763 1763 POTENTIAL.
FT CARBOHYD 1908 1908 POTENTIAL.
SO SEQUENCE 2039 AA; 223589 MM; 666F9033 CRC32;

Query Match 100.0%; Score 186; DB 1; Length 2039;
Best Local Similarity 100.0%; Pred. No. 1,8e-17;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CNPSSGGRKRVFELVGEPSICTSNDQVGIMSG 33
Db 195 CNPSSGGRKRVFELVGEPSICTSNDQVGIMSG 227

RESULT 2
CR2_HUMAN STANDARD; PRT; 1033 AA.
AC P20023;
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE COMPLEMENT RECEPTOR TYPE 2 PRECURSOR (CR2) (COMPLEMENT C3D RECEPTOR)
DE (EPSTEIN-BARR VIRUS RECEPTOR) (CD21 ANTIGEN).
GN CR2 OR C3DR.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89123277.
RA FUJISAKI A., HARLEY J.B., FRANK M.B., GRUNER B.A., FRAZIER B.,
RA HOLERS V.M.;

```





CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.  
CC -1- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR  
CC FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE  
CC ACTIVE SITE ON SCR3 (BY SIMILARITY).  
CC -1- PTM: THE SER/THR-RICH DOMAIN IS HEAVILY O-GLYCOSYLATED.  
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF DAF (DAF-2, SHOWN HERE, AND  
CC DAF-1) ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.  
CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.  
CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION  
CC (RCA) FAMILY.  
DR EMBL: S67775.6459390. -  
KW COMPLEMENT PATHWAY; PLASMA; GLYCOPROTEIN; MEMBRANE; REPEAT;  
KW ALTERNATIVE SPLICING; GPI-ANCHOR; SUSHI.  
FT NON\_TER 1 1  
FT CHAIN <1 312 COMPLEMENT DECAY-ACCELERATING FACTOR.  
FT PROPEP 313 340 REMOVED IN MATURE FORM (BY SIMILARITY).  
FT DOMAIN <1 243 4 X SUSHI (SCR) REPEATS.  
FT REPEAT <1 54 SUSHI 1.  
FT REPEAT 56 118 SUSHI 2.  
FT REPEAT 121 180 SUSHI 3.  
FT REPEAT 183 243 SUSHI 4.  
FT DOMAIN 246 315 SER/THR-RICH.  
FT DISULFID 24 53 BY SIMILARITY.  
FT DISULFID 57 104 BY SIMILARITY.  
FT DISULFID 88 117 BY SIMILARITY.  
FT DISULFID 122 163 BY SIMILARITY.  
FT DISULFID 149 179 BY SIMILARITY.  
FT DISULFID 184 226 BY SIMILARITY.  
FT DISULFID 212 242 BY SIMILARITY.  
FT CARBOHYD 54 54 POTENTIAL.  
FT CARBOHYD 107 107 POTENTIAL.  
FT LIPID 312 312 GPI-ANCHOR (BY SIMILARITY).  
SQ SEQUENCE 340 AA; 37180 MW; 5C4C18F8 CRC32;  
  
Query Match 44.9%; Score 83.5; DB 1; Length 340;  
Best Local Similarity 51.5%; Pred. No. 0.00029;  
Matches 17; Conservative 4; Mismatches 5; Indels 7; Gaps 2;  
  
OY 1 CNPSSGGRKVELGPEPISCTSDNDGIVMSG 33  
DB 212 CNKG-----YTMIGHSITCTVNDDE-GEWSG 237  
  
RESULT 4  
ID DAF\_HUMAN STANDARD; PRT; 381 AA.  
AC P08174; P09679;  
DT 01-AUG-1988 (REL. 08, CREATED)  
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE COMPLEMENT DECAY-ACCELERATING FACTOR PRECURSOR (CD55).  
GN DAF OR CD55  
OS HOMO SAPIENS (HUMAN).  
OC EUDAROTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC ETHIERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 87115845.  
RA CARAS I.W., DAVITZ M.A., RHEE L., WEDDELL G., MARTIN D.W. JR.,  
RA NUSSENZWEIG V.;  
RL NATURE 325:545-549(1987).  
RN [2]  
RP SEQUENCE OF 6-381 FROM N.A. (DAF-2 FORM).  
RX MEDLINE: 87175602.  
RA MEDOF M.E., LOBLIN D.M., HOLERS V.M., AYERS D.J., GETTY R.R.,  
RA LEKAM J.F., ATKINSON J.P., TYKOCINSKI M.L.;  
RL PROC. NATL. ACAD. SCI. U.S.A. 84:2007-2011(1987).  
RN [3]  
RP SEQUENCE OF 1-100 FROM N.A.  
RX MEDLINE: 91271256.  
RA EMOLONO U.K., RAVI L., MEDOF M.E.;  
RL PROC. NATL. ACAD. SCI. U.S.A. 88:4675-4679(1991).

RN [4]  
RP SEQUENCE OF 35-46.  
RC TISSUE-URINE;  
RX MEDLINE: 91291869.  
RA NAKANO Y., SUGITA Y., ISHIKAWA Y., CHOI N.-H., TOBE T., TOMITA M.;  
RL BIOCHIM. BIOPHYS. ACTA 1074:326-330(1991).  
RN [5]  
RP GPI-ANCHOR.  
RX MEDLINE: 91093238.  
RA MORAN P., RAAB H., KOHR W.J., CARAS I.W.;  
RL J. BIOL. CHEM. 266:1250-1257(1991).  
RN [6]  
FT DISULFIDE BONDS IN SUSHI DOMAINS.  
RX MEDLINE: 92305034.  
RA NAKANO Y., SUMIDA K., KIKUTA N., MIURA N.-H., TOBE T., TOMITA M.;  
RL BIOCHIM. BIOPHYS. ACTA 1116:235-240(1992).  
RN [7]  
RP FUNCTION AS A ECHOVIRUS RECEPTOR.  
RX MEDLINE: 95045399.  
RA WARD T., PIRKIN P.A., CLARKSON N.A., STONE D.M., MINOR P.D.,  
RA ALMOND J.W.;  
RL EMBO J. 13:5070-5074(1994).  
RN [8]  
RP VARIANT BLOOD GROUP DR(A-).  
RX MEDLINE: 94325573.  
RA LOBLIN D.M., MALLINSON G., POOLE J., REID M.E., THOMPSON E.S.,  
RA FERDAN B.R., TELEN M.J., ANSTEE D.J., TANNER M.J.A.;  
RL BLOOD 84:1276-1282(1994).  
CC -1- FUNCTION: THIS PROTEIN RECOGNIZES C4B AND C3B FRAGMENTS THAT  
CC CONDENSE WITH CELL-SURFACE HYDROXYL OR AMINO GROUPS WHEN NASCENT  
CC C4B AND C3B ARE LOCALLY GENERATED DURING C4 AND C3 ACTIVATION.  
CC INTERACTION OF DAF WITH CELL-ASSOCIATED C4B AND C3B POLYPEPTIDES  
CC INTERFERES WITH THEIR ABILITY TO CATALYZE THE CONVERSION OF C2 AND  
CC FACTOR B TO ENZYMATICALLY ACTIVE C2A AND Bb AND THEREBY PREVENTS  
CC THE FORMATION OF C4B2A AND C3BBB, THE AMPLIFICATION CONVERTASES OF  
CC THE COMPLEMENT CASCADE.  
CC -1- FUNCTION: ALSO ACTS AS THE RECEPTOR FOR ECHOVIRUS 7 AND RELATED  
CC VIRUSES (ECHOVIRUSES 13, 21, 29 AND 33).  
CC -1- SUBUNIT: MONOMER (MAJOR FORM) AND NON-DISULFIDE-LINKED, COVALENT  
CC HOMODIMER (MINOR FORM).  
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.  
CC -1- TISSUE SPECIFICITY: EXPRESSED ON THE PLASMA MEMBRANES OF ALL CELL  
CC TYPES THAT ARE IN INTIMATE CONTACT WITH PLASMA COMPLEMENT  
CC PROTEINS. IT IS ALSO FOUND ON THE SURFACES OF EPITHELIAL CELLS  
CC LINING EXTRACELLULAR COMPARTMENTS, AND VARIANTS OF THE MOLECULE  
CC ARE PRESENT IN BODY FLUIDS AND IN EXTRACELLULAR MATRIX.  
CC -1- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR  
CC FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE  
CC ACTIVE SITE ON SCR3 (BY SIMILARITY).  
CC -1- PTM: THE SER/THR-RICH DOMAIN IS HEAVILY O-GLYCOSYLATED.  
CC -1- POLYMORPHISM: DAF IS RESPONSIBLE FOR THE CROMER BLOOD GROUP  
CC SYSTEM. IT CONSISTS OF AT LEAST SEVEN HIGH-INCIDENCE (CR(A),  
CC TC(A), DR(A), ES(A), WES(B), WUC, AND IFC) AND LOW-INCIDENCE  
CC (TC(B), TC(C), AND WES(A)) ANTIGENS THAT RESIDE ON DAF. IN THE  
CC CROMER PHENOTYPES DR(A-) AND INAB THERE IS REDUCED OR ABSENT  
CC EXPRESSION OF DAF, RESPECTIVELY. IN THE CASE OF THE DR(A-)   
CC PHENOTYPE, A SINGLE NUCLEOTIDE SUBSTITUTION WITHIN EXON 5 ACCOUNTS  
CC FOR TWO CHANGES: A SIMPLE AMINO ACID SUBSTITUTION THAT IS THE  
CC BASIS OF THE ANTIGENIC VARIATION, AND AN ALTERNATIVE SPLICING  
CC EVENT THAT UNDERLIES THE DECREASED EXPRESSION OF DAF IN THIS  
CC PHENOTYPE.  
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF DAF (DAF-2, SHOWN HERE, AND  
CC DAF-1) ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.  
CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.  
CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION  
CC (RCA) FAMILY.  
CC -1- DATABASE: NAME-PROV: NOTE-CD guide CD55 entry;  
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd55.htm".  
DR EMBL: M31516; G181468; -  
DR EMBL: M30142; G181465; -  
DR EMBL: M15799; G181463; -  
DR EMBL: M64653; G181476; -  
DR EMBL: M64356; G181476; JOINED.



FT DISULFID 778 820 BY SIMILARITY.  
 FT DISULFID 806 833 BY SIMILARITY.  
 FT DISULFID 841 884 BY SIMILARITY.  
 FT DISULFID 870 897 BY SIMILARITY.  
 FT DISULFID 902 945 BY SIMILARITY.  
 FT DISULFID 931 958 BY SIMILARITY.  
 FT CARBOHYD 77 77 POTENTIAL.  
 FT CARBOHYD 113 113 POTENTIAL.  
 FT CARBOHYD 276 316 POTENTIAL.  
 FT CARBOHYD 316 316 POTENTIAL.  
 FT CARBOHYD 364 364 POTENTIAL.  
 FT CARBOHYD 380 380 POTENTIAL.  
 FT CARBOHYD 484 484 POTENTIAL.  
 FT CARBOHYD 527 527 POTENTIAL.  
 FT CARBOHYD 615 615 POTENTIAL.  
 FT CARBOHYD 639 639 POTENTIAL.  
 FT CARBOHYD 694 694 POTENTIAL.  
 FT CARBOHYD 754 754 POTENTIAL.  
 FT CARBOHYD 790 790 POTENTIAL.  
 FT CARBOHYD 813 813 POTENTIAL.  
 FT CARBOHYD 823 823 POTENTIAL.  
 FT CARBOHYD 851 851 POTENTIAL.  
 FT CARBOHYD 901 901 POTENTIAL.  
 FT CONFLICT 289 291 YGS -> EFR (IN REF. 4).  
 FT CONFLICT 306 306 S -> T (IN REF. 2).  
 FT CONFLICT 520 520 P -> A (IN REF. 2).  
 FT CONFLICT 562 563 MISSING (IN REF. 4).  
 SQ SEQUENCE 1025 AA; 112994 MW; DA043B62 CRC32;

Query Match 41.9%; Score 78; DB 1; Length 1025;  
 Best Local Similarity 42.4%; Pred. No. 0.0054;  
 Matches 14; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVFELVGEPSICTSNDQVGIWSG 33  
 DB 297 CDPSPKGVSTFLIGETINCTGSKRTGIWSG 329

RESULT 6  
 DAF1\_MOUSE STANDARD; PRT; 390 AA.  
 ID DAF1\_MOUSE  
 AC 061475; Q61397;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE COMPLEMENT DECAT-ACCELERATING FACTOR, GPI-ANCHORED PRECURSOR  
 DE (DAF-GPI).  
 GN DAF1.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;  
 RX MEDLINE: 95403982.  
 RA SPICER A.P., SELDIN M.F., GENDLER S.J.;  
 RL J. IMMUNOL. 155:3079-3091(1995).  
 RN [2]  
 RP SEQUENCE OF 7-390 FROM N.A.  
 RC STRAIN=BALB/C; TISSUE=SPLEEN;  
 RX MEDLINE: 96362213.  
 RA FUKUOKA Y., YASUI A., OKADA N., OKADA H.;  
 RL INT. IMMUNOL. 8:379-385(1996).  
 CC -1- FUNCTION: PROTECTION OF CELLS FROM COMPLEMENT-MEDIATED DAMAGE (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.  
 CC -1- TISSUE SPECIFICITY: BRAIN, SECRETORY EPITHELIA, SKELETAL MUSCLE, LIVER, TESTES, THYMUS, SPLEEN AND LYMPH NODE.  
 CC -1- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR THE FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE ACTIVE SITE ON SCRs (BY SIMILARITY).  
 CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.

CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION (RCA) FAMILY.  
 CC EMBL: L41366; G886335; -  
 DR EMBL: D63679; D1010476; -  
 DR MCD; MGI:104850; DAF1.  
 KW COMPLEMENT PATHWAY; GLYCOPROTEIN; MEMBRANE; REPEAT; GPI-ANCHOR;  
 KW SIGNAL; SUSHI.  
 FT CHAIN 1 34  
 FT SIGNAL 35 361 POTENTIAL.  
 FT PROPER 362 390 COMPLEMENT DECAT-ACCELERATING FACTOR,  
 FT DOMAIN 35 285 GPI-ANCHORED.  
 FT REPEAT 35 95 REMOVED IN MATURE FORM (BY SIMILARITY).  
 FT REPEAT 159 159 4 X SUSHI (SCR) REPEATS.  
 FT REPEAT 162 221 SUSHI 1.  
 FT REPEAT 224 285 SUSHI 2.  
 FT DOMAIN 288 364 SUSHI 3.  
 FT DISULFID 65 94 SUSHI 4.  
 FT DISULFID 98 145 SER/THR-RICH.  
 FT DISULFID 129 158 BY SIMILARITY.  
 FT DISULFID 163 204 BY SIMILARITY.  
 FT DISULFID 190 220 BY SIMILARITY.  
 FT DISULFID 225 267 BY SIMILARITY.  
 FT DISULFID 253 284 BY SIMILARITY.  
 FT CARBOHYD 187 187 POTENTIAL.  
 FT CARBOHYD 262 262 POTENTIAL.  
 FT LIPID 361 361 GPI-ANCHOR (BY SIMILARITY).  
 FT CONFLICT 7 7 P -> A (IN REF. 2).  
 FT CONFLICT 9 9 T -> A (IN REF. 2).  
 FT CONFLICT 83 83 E -> G (IN REF. 2).  
 FT CONFLICT 91 91 E -> K (IN REF. 2).  
 FT CONFLICT 135 135 E -> K (IN REF. 2).  
 FT CONFLICT 173 173 H -> L (IN REF. 2).  
 FT CONFLICT 180 180 I -> T (IN REF. 2).  
 SQ SEQUENCE 390 AA; 42618 MW; ADBBDB34 CRC32;

Query Match 36.0%; Score 67; DB 1; Length 390;  
 Best Local Similarity 58.3%; Pred. No. 0.062;  
 Matches 14; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

OY 9 KVELVGEPSICTSNDQVGIWS 32  
 DB 255 KGFILVGNASICTVSKSDVGWS 278

RESULT 7  
 C4BP\_MOUSE STANDARD; PRT; 469 AA.  
 ID C4BP\_MOUSE  
 AC P08607;  
 DT 01-AUG-1988 (REL. 08, CREATED)  
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE C4B-BINDING PROTEIN PRECURSOR (C4BP).  
 GN C4BPA OR C4BP.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 88024997.  
 RA KRISTENSEN T., OGATA R.T., REID K.B.M., TACK B.F.;  
 RL BIOCHEMISTRY 26:468-4674(1987).  
 CC -1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT ALSO ACCELERATES THE DEGRADATION OF THE C4Bc2a COMPLEX (C3 CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM AMYLOID P COMPONENT.  
 CC -1- SUBUNIT: HOMOPOLYMER; NOT COVALENTLY LINKED. MOUSE LACKS THE BETA CHAIN OF C4BP.  
 CC -1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.  
 CC -1- SIMILARITY: CONTAINS 6 SUSHI (SCR) REPEATS.





Query Match 32.5%; Score 60.5; DB 1; Length 377;  
Best Local Similarity 40.6%; Pred. No. 0.46;  
Matches 13; Conservative 5; Mismatches 9; Indels 5; Gaps 2;

QY 1 CNPSSGKRVFELVGEPSICTSNDVOYGIMS 32  
DB 191 CDP-APGDPFSLIGESTICYGDN---SVMS 217

## RESULT 10

CABP\_RAT  
ID CABP\_RAT STANDARD; PRT; 558 AA.  
AC 063514;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE CAB-BINDING PROTEIN ALPHA CHAIN PRECURSOR.  
GN CABPA.  
OS RATRUS NORVEGICUS (RAT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;  
RX MEDLINE: 97166082.  
RA HILLARP A., WIKLUND H., TERNER A., DAHLBACK B.;  
J. IMMUNOL. 158:1315-1323(1997).  
CC -1- FUNCTION: CABP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT ACTIVATION. IT BINDS AS A COPROCTOR TO C3B/C4B INACTIVATOR (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT ALSO ACCELERATES THE DEGRADATION OF THE C4B2A COMPLEX (C3 CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA CHAIN BINDS C4B. IT INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S AND WITH SERUM AMLOID P COMPONENT.  
CC -1- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS.  
CC -1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.  
DR EMBL: Z50051; G899380; -  
KM COMPLEMENT PATHWAY; PLASMA; GLYCOPROTEIN; REPEAT; SUSHI; SIGNAL.  
FT SIGNAL 1 13  
FT CHAIN 14 558  
FT DOMAIN 14 502  
FT REPEAT 14 73  
FT REPEAT 76 135  
FT REPEAT 138 200  
FT REPEAT 203 259  
FT REPEAT 262 325  
FT REPEAT 328 387  
FT REPEAT 389 444  
FT REPEAT 446 502  
FT DISULFID 15 60  
FT DISULFID 45 72  
FT DISULFID 77 118  
FT DISULFID 104 134  
FT DISULFID 139 182  
FT DISULFID 168 199  
FT DISULFID 204 246  
FT DISULFID 232 258  
FT DISULFID 263 312  
FT DISULFID 296 324  
FT DISULFID 328 351  
FT DISULFID 329 373  
FT DISULFID 336 386  
FT DISULFID 390 431  
FT DISULFID 417 443  
FT DISULFID 447 488  
FT DISULFID 474 501  
FT DISULFID 509 509  
FT DISULFID 521 521  
FT CARBOHYD 31 31  
FT CARBOHYD 177 177  
FT CARBOHYD 186 186  
FT CARBOHYD 469 469

FT CARBOHYD 491 491 POTENTIAL.  
SQ SEQUENCE 558 AA; 62266 MM; A08FE27A CRC32.

Query Match 32.3%; Score 60; DB 1; Length 558;  
Best Local Similarity 40.6%; Pred. No. 0.82;  
Matches 13; Conservative 4; Mismatches 9; Indels 6; Gaps 1;

QY 1 CNPSSGKRVFELVGEPSICTSNDVOYGIMS 32  
DB 168 CDP-----FTLGNASTICTVYNTKTVGWS 193

## RESULT 11

DAF2\_MOUSE  
ID DAF2\_MOUSE STANDARD; PRT; 407 AA.  
AC 061476;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE COMPLEMENT DECAY-ACCELERATING FACTOR, TRANSMEMBRANE PRECURSOR (DAF-TM).  
GN DAF2.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=TESTIS;  
RX MEDLINE: 95403982.  
RA SPICER A.P., SELDIN M.F., GENDLER S.J.;  
J. IMMUNOL. 155:3079-3091(1995).  
CC -1- FUNCTION: PROTECTION OF CELLS FROM COMPLEMENT-MEDIATED DAMAGE (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).  
CC -1- TISSUE SPECIFICITY: TESTES, SPLEEN AND LYMPH NODE.  
CC -1- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE ACTIVE SITE ON SCR3 (BY SIMILARITY).  
CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.  
CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION (RCA) FAMILY.  
DR EMBL: LA1365; G886338; -  
DR MGD: MGT:104849; DAF2.  
KM COMPLEMENT PATHWAY; GLYCOPROTEIN; REPEAT; SIGNAL; SUSHI;  
KW TRANSMEMBRANE.  
FT CHAIN 1 39  
FT SIGNAL 40 407  
FT DOMAIN 40 368  
FT TRANSMEM 369 388  
FT DOMAIN 390 407  
FT DOMAIN 40 290  
FT REPEAT 40 100  
FT REPEAT 102 164  
FT REPEAT 167 226  
FT REPEAT 229 290  
FT REPEAT 291 363  
FT DOMAIN 291 363  
FT DISULFID 70 99  
FT DISULFID 103 150  
FT DISULFID 134 163  
FT DISULFID 168 209  
FT DISULFID 195 225  
FT DISULFID 230 272  
FT DISULFID 258 289  
FT CARBOHYD 192 192  
FT CARBOHYD 267 267  
SQ SEQUENCE 407 AA; 44469 MM; BABF07E9 CRC32;

Query Match 31.7%; Score 59; DB 1; Length 407;  
Best Local Similarity 50.0%; Pred. No. 0.81;  
Matches 12; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

OY 9 KVFELGEPSTICTSNDVOGIMS 32  
| | | | : | | | : | | |  
Db 260 KGFLFGSTICTYKSDVGQMS 283

## RESULT 12

HIG\_DROME STANDARD: PRT: 958 AA.

AC 009101:

DT 01-NOV-1995 (REL. 32, CREATED)

DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)

DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)

DE LOCOMOTION-RELATED PROTEIN HIKARU GENKI PRECURSOR.

GN HIG.

OS DROSOPHILA MELANOGASTER (FRUIT FLY).

OC EURARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=HEAD;

RX MEDLINE: 93213498.

RA HOSHINO M., MATSUZAKI F., NABESHIMA Y.-I., HAMA C.;

RL NEURON 10:395-407(1993). IN THE DEVELOPMENT OF CNS FUNCTIONS INVOLVED

CC -1- FUNCTION: HAS A ROLE IN THE DEVELOPMENT OF CNS FUNCTIONS INVOLVED

CC IN LOCOMOTOR ACTIVITY.

CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.

CC -1- TISSUE SPECIFICITY: EXPRESSED IN PCC NEURONS AND NEUROBLASTS

CC IN THE PROCEPHALIC NEUROGENIC REGION IN THE CENTRAL NERVOUS

CC SYSTEM.

CC -1- DEVELOPMENTAL STAGE: MOST ABUNDANT DURING AND/OR AFTER NEURONAL

CC DIFFERENTIATION AND DURING CELL SPECIFICATION OR AXOGENESIS.

CC -1- ALTERNATIVE PRODUCTS: FOUR DIFFERENT ISOFORMS, TYPES 1-4, EXIST

CC DUE TO ALTERNATIVE SPLICING OF THE SAME GENE. THE SEQUENCE OF

CC TYPE 3 IS SHOWN HERE.

CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.

CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS

CC ONE C2-LIKE DOMAIN.

DR EMBL: D13884; G391663; -

DR EMBL: D13885; G391665; -

DR EMBL: D13886; G391667; -

DR EMBL: D13887; G391669; -

DR FLXBASE: FB900010114; h19.

KW GLYCOPROTEIN; ALTERNATIVE SPLICING; IMMUNOGLOBULIN FOLD; REPEAT;

KW SUSHI; SIGNAL.

FT CHAIN 1 31

FT DOMAIN 630 709

FT REPEAT 713 952

FT REPEAT 713 952

FT REPEAT 713 952

FT REPEAT 713 952

FT REPEAT 713 952

FT REPEAT 713 952

FT REPEAT 713 952

FT REPEAT 713 952

FT REPEAT 713 952

FT REPEAT 713 952

FT REPEAT 713 952

FT REPEAT 713 952

FT REPEAT 713 952

FT REPEAT 713 952

FT REPEAT 713 952

FT REPEAT 713 952

FT REPEAT 713 952

FT REPEAT 713 952

FT REPEAT 713 952

FT REPEAT 713 952

FT REPEAT 713 952

FT REPEAT 713 952

FT REPEAT 713 952

Query Match 30.6%; Score 57; DB 1; Length 958;

Best Local Similarity 46.9%; Pred. No. 3.8;  
Matches 15; Conservative 0; Mismatches 7; Indels 10; Gaps 2;

OY 1 CNPGRKRVFELGEPSTICTSNDVOGIMS 32  
| | | | : | | | : | | |  
Db 863 CNDG-----HSLVGESSICTEN---GQWS 884

## RESULT 13

C4BP\_BOVIN STANDARD: PRT: 610 AA.

AC 028065;

DT 01-NOV-1997 (REL. 35, CREATED)

DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

DE C4B-BINDING PROTEIN ALPHA CHAIN PRECURSOR.

GN C4BPA.

OS BOS TAURUS (BOVINE).

OC EURARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; ARTIODACTYLA.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-LIVER;

RX MEDLINE: 95015909.

RA HILLARP A., THERN A., DAHLBACH B.;

RL J. IMMUNOL. 153:4190-4199(1994).

CC -1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT

CC ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR

CC (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT

CC ALSO ACCELERATES THE DEGRADATION OF THE C4B/C2A COMPLEX (C3

CC CONVERTASE) BY DISOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA

CC CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM AMYLOID P COMPONENT.

CC -1- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS

CC (BY SIMILARITY).

CC -1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.

CC -1- SIMILARITY: CONTAINS 8 SUSHI (SCR) REPEATS.

DR EMBL: Z31693; G469118; -

KW COMPLEMENT PATHWAY; PLASMA; GLYCOPROTEIN; REPEAT; SUSHI; SIGNAL.

FT CHAIN 1 48

FT DOMAIN 49 610

FT REPEAT 49 542

FT REPEAT 49 542

FT REPEAT 49 542

FT REPEAT 49 542

FT REPEAT 49 542

FT REPEAT 49 542

FT REPEAT 49 542

FT REPEAT 49 542

FT REPEAT 49 542

FT REPEAT 49 542

FT REPEAT 49 542

FT REPEAT 49 542

FT REPEAT 49 542

FT REPEAT 49 542

FT REPEAT 49 542

FT REPEAT 49 542

FT REPEAT 49 542

FT REPEAT 49 542

FT REPEAT 49 542

FT REPEAT 49 542

FT REPEAT 49 542

FT REPEAT 49 542

FT REPEAT 49 542

FT REPEAT 49 542

FT REPEAT 49 542

FT REPEAT 49 542

FT REPEAT 49 542

FT REPEAT 49 542

FT REPEAT 49 542

Query Match 610 AA; 68886 MW; 39814F78 CRC32;





FT  
FT VARSPLIC 459 507 TCS).  
THYKVDSEFAGASNHMLADIAKEDLRDRFSNAONISLLQ  
FT VLGAQOTQ -> ANRRHRRTKPNMLYVIT (IN GDA-TCS  
AND GDAB-TCS).  
FT  
SQ SEQUENCE 507 AA; 55263 MM; EF019A63 CRC32;

Query Match 28.2%; Score 52.5; DB 1; Length 507;  
Best Local Similarity 37.5%; Pred. No. 7.9;  
Matches 12; Conservative 4; Mismatches 9; Indels 7; Gaps 2;

OY 1 CNPSSGGRKVELVEPSIYCTSDNDQVIGWS 32  
1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1  
Db 189 CDPG-----YRLTGEASAFCMKGNVAVG-WS 213

RESULT 15  
HEXC\_BOMMO STANDARD; PRT; 596 AA.  
ID HEXC\_BOMMO  
AC P49010;  
DT 01-FEB-1996 (REL. 33, CREATED)  
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
DE CHITOOLOGOSACCHARIDOLYTIC BETA-N-ACETYLGLUCOSAMINIDASE PRECURSOR  
DE (EC 3.2.1.52) (BETA-GLUCNACASE) (BETA-HEXOSAMINIDASE) (BETA-N-  
ACETYLHEXOSAMINIDASE).  
OS BOMBYX MORI (SILK MOTH).  
OS EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; LEPIDOPTERA.  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN-KINSHU X SHOWA;  
RX MEDLINE; 95218264.  
RA NAGAMATSU Y., YANAGISAWA I., KIMOTO M., OKAMOTO E., KOGA D.;  
RL BIOSCI. BIOTECHNOL. BIOCHEM. 59:219-225(1995).  
CC -!- FUNCTION: ACTIVE DURING METAMORPHOSIS TO DEGRADE CHITIN  
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL NON-REDUCING N-ACETYL-  
CC -!- D-HEXOSAMINE RESIDUES IN N-ACETYL-BETA-D-HEXOSAMINIDES.  
CC -!- SIMILARITY: BELONGS TO FAMILY 20 OF GLYCOSYL HYDROLASES.  
DR EMBL; S77548; G998377; .  
KM HYDROLASE; GLYCOSIDASE; SIGNAL; GLYCOPROTEIN.  
FT SIGNAL 1 23 POTENTIAL.  
FT CHAIN 24 596 CHITOOLOGOSACCHARIDOLYTIC BETA-N-  
FT ACETYLGLUCOSAMINIDASE.  
FT CARBOHYD 166 166 POTENTIAL.  
FT CARBOHYD 264 264 POTENTIAL.  
FT CARBOHYD 377 377 POTENTIAL.  
SQ SEQUENCE 596 AA; 68213 MM; D56ECAEC CRC32;

Query Match 28.2%; Score 52.5; DB 1; Length 596;  
Best Local Similarity 35.1%; Pred. No. 9.4;  
Matches 13; Conservative 6; Mismatches 11; Indels 7; Gaps 2;

OY 1 CNPSSGGRKVELVEPSIYCTSDNDQ-----VGIWS 32  
1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1  
Db 493 CSPYIGGQKYG--NSPAVMALSYRQILGGEVALMS 527

Search completed: July 18, 1999, 11:37:49  
Job time: 1748 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 05:11:05 ; Search time 63.5 seconds  
(without alignments)  
28.671 Million cell updates/sec

Title: US-09-142-043-1

Perfect score: 186  
Sequence: 1 CNPGSGCRKRVFELVGEPSICYTSNDPOVGIMSG 33

Scoring table: BLOSUM62

Searched: 180763 segs, 55169189 residues

Database :

SPREMBL\_8.\*  
1: sp\_fungi.\*  
2: sp\_human.\*  
3: sp\_invertebrate.\*  
4: sp\_mammal.\*  
5: sp\_mhc.\*  
6: sp\_organelle.\*  
7: sp\_phage.\*  
8: sp\_plant.\*  
9: sp\_bacteria.\*  
10: sp\_rodent.\*  
11: sp\_virus.\*  
12: sp\_vertibrate.\*  
13: sp\_unclassified.\*  
14: sp\_archaea.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	186	100.0	2039	2	016745	016745 homo sapien
2	186	100.0	2489	2	016744	016744 homo sapien
3	186	100.0	661	4	029531	029531 pan troglod
4	186	100.0	2014	4	029530	029530 pan troglod
5	169	90.9	315	4	028770	028770 papio hamad
6	169	90.9	1911	4	029528	029528 papio hamad
7	167	89.8	522	4	028797	028797 pan troglod
8	159	85.5	479	2	014079	014079 homo sapien
9	156	83.9	522	4	028769	028769 papio hamad
10	129	69.4	483	10	064735	064735 mus musculu
11	127	68.3	417	10	035520	035520 rattus norv
12	127	68.3	497	10	063612	063612 rattus norv
13	127	68.3	559	10	063135	063135 rattus norv
14	116	62.4	679	10	099254	099254 mus musculu
15	113	60.8	89	10	063129	063129 rattus norv
16	98	52.7	1032	2	013866	013866 homo sapien
17	98	52.7	1087	2	014212	014212 homo sapien
18	85	45.7	1045	4	046545	046545 ovis aries
19	81.5	43.8	347	2	P78361	P78361 homo sapien
20	71	38.2	330	10	062479	062479 mus musculu
21	67.5	36.3	336	4	062834	062834 saguinus oe
22	67.5	36.3	314	4	062835	062835 saguinus oe
23	67.5	36.3	378	4	062837	062837 saguinus oe
24	67.5	36.3	377	4	062838	062838 saguinus oe
25	67.5	36.3	222	4	019122	019122 callimico g
26	67.5	36.3	222	4	019124	019124 saguinus my
27	67.5	36.3	222	4	019128	019128 pithecia pl
28	67	36.0	390	10	P97732	P97732 mus musculu
29	62.5	33.6	222	4	019120	019120 aotus trivl

## ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	2039 AA.	
Q16745	Q16745				
ID	Q16745				
AC	Q16745				
DT	01-NOV-1996 (TREMBLREL. 01, CREATED)				
DT	01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)				
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)				
DE	COMPLEMENT RECEPTOR 1.				
OS	HOMO SAPIENS (HUMAN).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;				
OC	CATARRHINI; HOMINIDAE; HOMO.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE: 94065175.				
RA	VIR D.P., WONG W.W.;				
RT	*Structure of the gene for the F allele of complement receptor type 1				
RT	and sequence of the coding region unique to the S allele.*;				
RL	J. IMMUNOL. 151:6214-6224(1993).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	VIR D.P., WONG W.W.;				
RL	SUBMITTED (JUN-1993) TO EMBL/GENBANK/DBJ DATA BANKS.				
DR	EMBL: L17399; G306680; JOINED.				019125 saimiri sci
DR	EMBL: L17419; G306680; JOINED.				015429 homo sapien
DR	EMBL: L17419; G306680; JOINED.				019123 callithrix
DR	EMBL: L17421; G306680; JOINED.				041927 murine herp
DR	EMBL: L17422; G306680; JOINED.				019121 papio hamad
DR	EMBL: L17423; G306680; JOINED.				019126 macaca fasc
DR	EMBL: L17381; G306680; JOINED.				060736 mus musculu
DR	EMBL: L17382; G306680; JOINED.				P79138 cercopithec
DR	EMBL: L17393; G306680; JOINED.				002839 sus scrofa
DR	EMBL: L17395; G306680; JOINED.				022328 caenorhabd
DR	EMBL: L17396; G306680; JOINED.				008569 cavia porce
DR	EMBL: L17397; G306680; JOINED.				091746 xenopus lae
DR	EMBL: L17398; G306680; JOINED.				P79927 xenopus lae
DR	EMBL: L17398; G306680; JOINED.				019261 caenorhabd
DR	EMBL: L17418; G306680; JOINED.				093967 caenorhabd
DR	EMBL: L17390; G306680; JOINED.				
DR	EMBL: L17400; G306680; JOINED.				
DR	EMBL: L17416; G306680; JOINED.				
DR	EMBL: L17417; G306680; JOINED.				
DR	EMBL: L17401; G306680; JOINED.				
DR	EMBL: L17403; G306680; JOINED.				
DR	EMBL: L17403; G306680; JOINED.				
DR	EMBL: L17404; G306680; JOINED.				
DR	EMBL: L17405; G306680; JOINED.				
DR	EMBL: L17406; G306680; JOINED.				
DR	EMBL: L17407; G306680; JOINED.				
DR	EMBL: L17408; G306680; JOINED.				
DR	EMBL: L17410; G306680; JOINED.				
DR	EMBL: L17411; G306680; JOINED.				

DR EMBL: L17412: G306680; JOINED.  
DR EMBL: L17413: G306680; JOINED.  
DR EMBL: L17414: G306680; JOINED.  
DR EMBL: L17415: G306680; JOINED.  
DR PFAM: PF00084: sushi; 30.  
SQ SEQUENCE 2039 AA: 223603 MW: A4972215 CRC32;

Query Match 100.0%; Score 186; DB 2; Length 2039;  
Best Local Similarity 100.0%; Pred. No. 1,5e-17;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVFELVGEPSICTSNDQVIGMSG 33  
Db 195 CNPSSGGRKVFELVGEPSICTSNDQVIGMSG 227  
|||||

RESULT 2  
ID Q16744 PRELIMINARY; PRT: 2489 AA.  
AC Q16744;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE COMPLEMENT RECEPTOR 1.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; HOMO.  
RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE: 94065175.  
RA VIK D.P., WONG W.W.;  
RT "Structure of the gene for the F allele of complement receptor type 1  
and structure of the coding region unique to the S allele."  
RL J. IMMUNOL. 151:6214-6224(1993).

RN [2]  
RP SEQUENCE FROM N.A.  
RX VIK D.P., WONG W.W.;  
RL SUBMITTED (JUN-1993) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL: L17399: G451303; JOINED.  
DR EMBL: L17409: G451303; JOINED.  
DR EMBL: L17419: G451303; JOINED.  
DR EMBL: L17420: G451303; JOINED.  
DR EMBL: L17421: G451303; JOINED.  
DR EMBL: L17422: G451303; JOINED.  
DR EMBL: L17423: G451303; JOINED.  
DR EMBL: L17424: G451303; JOINED.  
DR EMBL: L17425: G451303; JOINED.  
DR EMBL: L17426: G451303; JOINED.  
DR EMBL: L17427: G451303; JOINED.  
DR EMBL: L17428: G451303; JOINED.  
DR EMBL: L17429: G451303; JOINED.  
DR EMBL: L17430: G451303; JOINED.  
DR EMBL: L17391: G451303; JOINED.  
DR EMBL: L17392: G451303; JOINED.  
DR EMBL: L17393: G451303; JOINED.  
DR EMBL: L17394: G451303; JOINED.  
DR EMBL: L17395: G451303; JOINED.  
DR EMBL: L17396: G451303; JOINED.  
DR EMBL: L17397: G451303; JOINED.  
DR EMBL: L17398: G451303; JOINED.  
DR EMBL: L17400: G451303; JOINED.  
DR EMBL: L17401: G451303; JOINED.  
DR EMBL: L17402: G451303; JOINED.  
DR EMBL: L17403: G451303; JOINED.  
DR EMBL: L17405: G451303; JOINED.  
DR EMBL: L17406: G451303; JOINED.  
DR EMBL: L17407: G451303; JOINED.  
DR EMBL: L17408: G451303; JOINED.

DR EMBL: L17410: G451303; JOINED.  
DR EMBL: L17411: G451303; JOINED.  
DR EMBL: L17412: G451303; JOINED.  
DR EMBL: L17413: G451303; JOINED.  
DR EMBL: L17414: G451303; JOINED.  
DR EMBL: L17415: G451303; JOINED.  
DR PFAM: PF00084: sushi; 37.  
SQ SEQUENCE 2489 AA: 272846 MW: 586986F9 CRC32;

Query Match 100.0%; Score 186; DB 2; Length 2489;  
Best Local Similarity 100.0%; Pred. No. 1,8e-17;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVFELVGEPSICTSNDQVIGMSG 33  
Db 195 CNPSSGGRKVFELVGEPSICTSNDQVIGMSG 227  
|||||

RESULT 3  
ID Q29531 PRELIMINARY; PRT: 661 AA.  
AC Q29531;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE COMPLEMENT RECEPTOR 1 PRECURSOR (FRAGMENT).  
GN CRI.  
OS PAN TROGLODYTES (CHIMPANZEE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; PAN.  
RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE: 94292799.  
RA BIRNIGHAM D.J., SHEN X.P., HOURCADE D., NICKELLS M.W., ATKINSON J.P.;  
RT "Primary sequence of an alternatively spliced form of CRI. Candidate  
RT for the 75,000 M(r) complement receptor expressed on chimpanzee  
erythrocytes."  
RL J. IMMUNOL. 153:691-700(1994).

DR EMBL: L24921: G557727; -  
DR PFAM: PF00084: sushi; 9.  
KW SIGNAL; ALTERNATIVE SPLICING.  
FT NON\_TER 1  
FT SIGNAL 1  
FT CHAIN 16  
FT CHAIN 17  
FT NON\_TER 661  
SQ SEQUENCE 661 AA: 72966 MW: 9D78E262 CRC32;

Query Match 100.0%; Score 186; DB 4; Length 661;  
Best Local Similarity 100.0%; Pred. No. 4,6e-18;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVFELVGEPSICTSNDQVIGMSG 33  
Db 170 CNPSSGGRKVFELVGEPSICTSNDQVIGMSG 202  
|||||

RESULT 4  
ID Q29530 PRELIMINARY; PRT: 2014 AA.  
AC Q29530;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE COMPLEMENT RECEPTOR 1 (FRAGMENT).  
GN CRI.  
OS PAN TROGLODYTES (CHIMPANZEE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; PAN.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 94292799.

RA BIRMINGHAM D.J., SHEN X.P., HOURCADE D., NICKELLS M.W., ATKINSON J.P.;  
RT "Primary sequence of an alternatively spliced form of CRI. Candidate  
for the 75,000 M(r) complement receptor expressed on chimpanzee  
erythrocytes."  
RL J. IMMUNOL. 153:691-700(1994).  
DR EMBL: L24920: G557725; -.  
DR PFAM: PF00084; sushi; 30.  
FT NON\_TER 1  
FT NON\_TER 2014 2014  
SQ SEQUENCE 2014 AA: 221280 MW: 1370CDD8 CRC32;

Query Match 100.0%; Score 186; DB 4; Length 2014;  
Best Local Similarity 100.0%; Pred. No. 1.4e-17;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVELGEPISICTSNDQVGIMSG 33  
Db 170 CNPSSGGRKVELGEPISICTSNDQVGIMSG 202

RESULT 5  
Q28770 PRELIMINARY; PRT; 315 AA.  
AC Q28770;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE COMPLEMENT RECEPTOR (FRAGMENT).  
OS PAPIO HAMADRYAS (HAMADRYAS BABOON).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; PAPIO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BONE MARROW.  
RA BIRMINGHAM D.J., LOGAR C.M., SHEN X.P., CHEN W.;  
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: L77978; G1301611; -.  
DR PFAM: PF00084; sushi; 5.  
FT NON\_TER 1  
FT NON\_TER 315  
SQ SEQUENCE 315 AA: 35172 MW: 5776DB8A CRC32;

Query Match 90.9%; Score 169; DB 4; Length 315;  
Best Local Similarity 90.9%; Pred. No. 5e-16;  
Matches 30; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVELGEPISICTSNDQVGIMSG 33  
Db 154 CNPSSGGRKVELGEPISICTSNDQVGIMSG 186

RESULT 6  
Q29528 PRELIMINARY; PRT; 1911 AA.  
AC Q29528;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE COMPLEMENT RECEPTOR 1 (FRAGMENT).  
OS PAPIO HAMADRYAS (HAMADRYAS BABOON).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; PAPIO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA CLEMENZA L., SUBRAMANIAN B.V., NICKELLS M.W., HOURCADE D.E.,  
ATKINSON J.P.;  
RL SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: L39791; G662829; -.  
DR PFAM: PF00084; sushi; 29.  
FT NON\_TER 1911 1911

SQ SEQUENCE 1911 AA: 210173 MW: 03D640C2 CRC32;

Query Match 90.9%; Score 169; DB 4; Length 1911;  
Best Local Similarity 90.9%; Pred. No. 3.2e-15;  
Matches 30; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVELGEPISICTSNDQVGIMSG 33  
Db 82 CNPSSGGRKVELGEPISICTSNDQVGIMSG 114

RESULT 7  
Q28797 PRELIMINARY; PRT; 522 AA.  
AC Q28797;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE UNKNOWN PROTEIN (FRAGMENT).  
OS PAN TROGLODYTES (CHIMPANZEE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; PAN.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA BIRMINGHAM D.J., SHEN X.P., HOURCADE D., NICKELLS M.W., ATKINSON J.P.;  
RL MEDLINE: 94292799.  
RT "Primary sequence of an alternatively spliced form of CRI. Candidate  
for the 75,000 M(r) complement receptor expressed on chimpanzee  
erythrocytes."  
RL J. IMMUNOL. 153:691-700(1994).  
DR EMBL: L24922: G557729; -.  
DR PFAM: PF00084; sushi; 8.  
FT NON\_TER 1  
SQ SEQUENCE 522 AA: 57304 MW: 21A32E12 CRC32;

Query Match 89.8%; Score 167; DB 4; Length 522;  
Best Local Similarity 90.9%; Pred. No. 1.6e-15;  
Matches 30; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVELGEPISICTSNDQVGIMSG 33  
Db 137 CNPSSGGRKVELGEPISICTSKDDQVGIMSG 169

RESULT 8  
Q14079 PRELIMINARY; PRT; 479 AA.  
AC Q14079;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE COMPLEMENT RECEPTOR 1-LIKE PROTEIN (CRL1) (FRAGMENT).  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE: 90110163.  
RA HOURCADE D., MIESNER D.R., BEE C., ZEIDES W., ATKINSON J.P.;  
RT "Duplication and divergence of the amino-terminal coding region of the  
complement receptor 1 (CRI) gene. An example of concerted (horizontal)  
evolution within a gene."  
RT J. BIOL. CHEM. 265:974-980(1990).  
RL EMBL: M31231; G563326; JOINED.  
DR EMBL: M31231; G563326; JOINED.  
DR EMBL: M31232; G563326; JOINED.  
DR EMBL: M31233; G563326; JOINED.  
DR EMBL: M31237; G563326; -.  
DR EMBL: M31230; G563326; JOINED.  
DR EMBL: M31234; G563326; JOINED.  
DR EMBL: X14359; G563326; JOINED.  
DR EMBL: X14360; G563326; JOINED.

DR EMBL; M31235; G5633326; JOINED.  
DR EMBL; M31236; G5633326; JOINED.  
DR PFAM; PF00084; sushi; 7.  
FT NON\_TER 479 479  
SQ SEQUENCE 479 AA; 52533 MW; EC994E0D CRC32;

Query Match 85.5%; Score 159; DB 2; Length 479;  
Best Local Similarity 87.9%; Pred. No. 1.9e-14;  
Matches 29; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CNPSSGGRKVELGEPISYCTSDNDQVIGWSG 33  
DB 195 CNLGSGRKVELGEPISYCTSDNDQVIGWSG 227

RESULT 9  
Q28769 PRELIMINARY; PRT; 522 AA.

AC Q28769;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE COMPLEMENT RECEPTOR (FRAGMENT).  
OS PAPIO HAMADRYAS (HAMADRYAS BABOON).  
OC EUDAROTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; PAPIO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BONE MARROW;  
RA BIRMINGHAM D.J.; LOGAR C.M., SHEN X.P., CHEN W.;  
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; L77977; G3301609; -  
DR PFAM; PF00084; sushi; 7.  
FT NON\_TER 1 1  
SQ SEQUENCE 522 AA; 56626 MW; 7862072C CRC32;

Query Match 83.9%; Score 156; DB 4; Length 522;  
Best Local Similarity 84.8%; Pred. No. 5.5e-14;  
Matches 28; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 CNPSSGGRKVELGEPISYCTSDNDQVIGWSG 33  
DB 200 CNLGSGRKVELGEPISYCTSDNDQVIGWSG 232

RESULT 10  
Q64735 PRELIMINARY; PRT; 483 AA.

AC Q64735;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE COMPLEMENT RECEPTOR RELATED PROTEIN.  
GN CRRY.  
OS MUS MUSCULUS (MOUSE).  
OC EUDAROTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
OC SCUROGNATHI; MURIDAE; MURINAE; MUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BALB/C;  
RX MEDLINE; 90171600.  
RA PAUL M.S.; AGEISTER-SHAM M., CEPER K., MILLER M.D., WEIS J.H.;  
RT "The murine complement receptor gene family. III. The genomic and  
transcriptional complexity of the Crry and Crry-ps genes.";  
RL J. IMMUNOL. 144:1988-1996(1990).  
DR EMBL; M34166; G387133; JOINED.  
DR EMBL; M34167; G387133; JOINED.  
DR EMBL; M34168; G387133; JOINED.  
DR EMBL; M34169; G387133; JOINED.  
DR EMBL; M34170; G387133; JOINED.  
DR EMBL; M34171; G387133; JOINED.

DR EMBL; M34172; G387133; JOINED.  
DR EMBL; M34173; G387132; -  
DR EMBL; M34173; G387133; -  
DR EMBL; M34164; G387133; JOINED.  
DR EMBL; M34165; G387133; JOINED.  
DR EMBL; M34164; G387132; JOINED.  
DR EMBL; M34170; G387132; JOINED.  
DR EMBL; M34171; G387132; JOINED.  
DR EMBL; M34165; G387132; JOINED.  
DR EMBL; M34166; G387132; JOINED.  
DR EMBL; M34167; G387132; JOINED.  
DR EMBL; M34168; G387132; JOINED.  
DR EMBL; M34169; G387132; JOINED.  
DR MGD; MGI:88513; CRRY.  
DR PFAM; PF00084; sushi; 7.  
FT VARIANT 41 83 MISSING (TN LIVER ISOFORM).  
SQ SEQUENCE 483 AA; 53762 MW; 70C73A80 CRC32;

Query Match 69.4%; Score 129; DB 10; Length 483;  
Best Local Similarity 60.6%; Pred. No. 3e-10;  
Matches 20; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 1 CNPSSGGRKVELGEPISYCTSDNDQVIGWSG 33  
DB 237 CNTDARGKALENLVGEPSYCTSDNGEIGWSG 269

RESULT 11  
Q35520 PRELIMINARY; PRT; 417 AA.

AC Q35520;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE 512 ANTIGEN (FRAGMENT).  
OS RATTUS NORVEGICUS (RAT).  
OC EUDAROTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
OC SCUROGNATHI; MURIDAE; MURINAE; RATTUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-LIVER;  
RA DOIH N., SAKURADA C., NONAKA M., OKADA N., OKADA H.;  
RL SUBMITTED (NOV-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; D42116; D1023417; -  
DR PFAM; PF00084; sushi; 6.  
FT NON\_TER 417 417  
SQ SEQUENCE 417 AA; 45951 MW; A6F91984 CRC32;

Query Match 68.3%; Score 127; DB 10; Length 417;  
Best Local Similarity 66.7%; Pred. No. 4.8e-10;  
Matches 22; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 CNPSSGGRKVELGEPISYCTSDNDQVIGWSG 33  
DB 52 CNTDARGKALENLVGEPSYCTSDNGEIGWSG 84

RESULT 12  
Q63612 PRELIMINARY; PRT; 497 AA.  
AC Q63612;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE 512 ANTIGEN PRECURSOR.  
OS RATTUS NORVEGICUS (RAT).  
OC EUDAROTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
OC SCUROGNATHI; MURIDAE; MURINAE; RATTUS.  
RN [1]  
RP SEQUENCE FROM N.A.



DR PFAM; PF00084; sushi; 1.  
 KW TRANSMEMBRANE.  
 FT NON\_TER 1  
 FT NON\_TER 89  
 FT NON\_TER 89  
 SQ SEQUENCE 89 AA; 9756 MW; 9FBA44EB CRC32;

Query Match 60.8%; Score 113; DB 10; Length 89;  
 Best Local Similarity 71.4%; Pred. No. 8.9e-09;  
 Matches 20; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVEFYGEPSIYCTSDNDY 28  
 11 : 1:1 11111111111111111111  
 Db 62 CNTDARGRKLEFNLVGEPSIHCTSDNDY 89

Search completed: July 18, 1999, 06:02:48  
 Job time: 3103 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 00:42:52 ; Search time 58.51 Seconds  
(without alignments)  
2.765 Million cell updates/sec

Title: US-09-142-043-2  
Perfect score: 39  
Sequence: 1 KKKKKRKA 8

Scoring table: BLOSUM62

Searched: 162890 seqs, 20225328 residues

Database: A\_Geneseq\_34:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	100.0	16	R65867	Antideoxyribonucle
2	39	100.0	17	R66010	Peptide which bind
3	36	92.3	116	W03642	Human cannabinoi
4	36	92.3	1040	W24559	Presentin-interac
5	36	92.3	756	W60664	Human ALAKA protei
6	35	89.7	783	P94265	Sequence of APh36.
7	35	89.7	744	P81006	Alpha-2-plasmin in
8	35	89.7	175	R08261	Peptide antigenic
9	35	89.7	12	R24220	Fragment of tenasc
10	35	89.7	12	P71165	Peptide which inh
11	35	89.7	40	R11512	Anti-ATLA antibody
12	35	89.7	29	R11513	Anti-ATLA antibody
13	35	89.7	34	R11514	Anti-ATLA antibody
14	35	89.7	898	P61030	Entire coded sequ
15	35	89.7	899	P61056	Translation of pla
16	35	89.7	898	P61082	Complete translati
17	35	89.7	317	R21750	C-terminal sequenc
18	35	89.7	242	R22390	Antigen ac-Ib. Vac
19	35	89.7	20	P20159	Sequence of lysine
20	35	89.7	152	R24045	Sequence of a pept
21	35	89.7	1167	R31376	Human flk-2. Nucle
22	35	89.7	474	R32587	P.falciparum anti
23	35	89.7	474	R32587	P.falciparum anti
24	35	89.7	36	R34595	Sequence of endoso
25	35	89.7	36	R34603	Sequence of endoso
26	35	89.7	36	R34604	Sequence of endoso
27	35	89.7	1093	R56979	Human myotonic dys
28	35	89.7	1093	R56979	Human myotonic dys
29	35	89.7	1189	R56496	TAIN-binding prote
30	35	89.7	7	R62111	Hydrophilic, basic
31	35	89.7	412	R63789	Aspergillus acula
32	35	89.7	11	R60841	Polyoctine octa-GXL
33	35	89.7	16	R66023	Pair 1 Pep 1 immun
34	35	89.7	16	R66024	Pair 1 Pep 2 immun
35	35	89.7	16	R66025	Pair 2 Pep 3 immun
36	35	89.7	16	R66026	Pair 2 Pep 4 immun
37	35	89.7	16	R66027	Pair 3 Pep 5 immun
38	35	89.7	16	R66028	Pair 3 Pep 6 immun
39	35	89.7	182	R62619	Deduced sequence o
40	35	89.7	418	R63676	Mitosis-associated
41	35	89.7	10	R71774	Peptide neutralisi
42	35	89.7	1235	R65158	Potassium ion chan
43	35	89.7	25	R86385	Gene therapy synth

ALIGNMENTS

44	35	89.7	473	1	R86502	Protein sequence f
45	35	89.7	20	1	W08403	Nucleic acid conde
RESULT 1						
R65867	standard; peptide: 16 AA.					
AC	R65867;					
DT	26-JUN-1995 (first entry)					
DE	Antideoxyribonucleic acid antibody adsorbent peptide.					
KW	Antideoxyribonucleic acid adsorbent peptide; kidney inflammation;					
OS	systemic lupus erythematosus.					
PN	Synthetic.					
PD	J062619411-A.					
PF	20-SEP-1994.					
PR	13-OCT-1993; 256111.					
PA	18-JAN-1993; JP-006100.					
DR	(KURS ) KURARAY CO LTD.					
PT	WPI: 94-337461/42.					
PS	Adsorbents of antideoxyribonucleic - comprising a peptide.					
PT	Immobilised on a carrier.					
PS	Claim 1; Page 15; 16pp; Japanese.					
CC	R65841-R65879 are antideoxyribonucleic (ADN) acid adsorbent peptides					
CC	derived from the highly generic peptide in the specification. When these					
CC	peptides are immobilised on a carrier (either cellulose, porous					
CC	glass, polyvinyl alcohol or polyacrylamide) they can be used to					
CC	remove ADN antibodies, which can result in kidney inflammation and					
CC	systemic lupus erythematosus.					
SQ	Sequence 16 AA:					

Query Match	100.0%;	Score 39;	DB 1;	Length 16;
Best Local Similarity	100.0%;	Pred. No. 0.76;		
Matches	8;	Conservative	0;	Mismatches 0;
Indels	0;	Gaps	0;	
RESULT 2				
R66010	standard; peptide: 17 AA.			
AC	R66010;			
DT	30-JUN-1995 (first entry)			
DE	Peptide which binds immune complexes.			
KW	Immune complex; immunoabsorbent.			
OS	Synthetic.			
PN	J06263795-A.			
PD	20-SEP-1994.			
PF	30-SEP-1993; 245367.			
PR	30-SEP-1992; JP-261821.			
PA	18-JAN-1993; JP-006099.			
DR	(KURS ) KURARAY CO LTD.			
PT	WPI: 94-338300/42.			
PS	Peptide(s) capable of binding to immune complex, and adsorbents			
PT	contg them - useful for removal of immune complexes from body			
PS	fluid and for treatment of associated diseases.			
CC	Example 12; Page 6; 15pp; Japanese.			
CC	The invention relates to new peptides of formula H-X-A-Y-Z in which			
CC	A is a peptide fragment of 6-12 amino acids containing the sequence			
CC	Ala-B-Glu-Ile-Leu, where B and C are Trp, Tyr and/or Phe; X and Y			
CC	are each a single bond, or an amino acid or peptide composed of 2-10			
CC	amino acids selected from Asp, Glu, Arg, Lys and His, provided at			
CC	least one of X and Y is not a bond; and Z is OH or amide.			
CC	The peptides can bind to immune complexes. They can be fixed to a			
CC	support and used for selective removal of immune complexes from body			
CC	fluids without removal of useful components. They can thus be used for			
CC	treating diseases involving the complexes. In Examples, the			
CC	activity of the peptides is compared to that of reference peptides			

CC in which neither X nor Y is present.  
CC The present sequence is a specific example of the new peptides.  
SQ Sequence 17 AA;

Query Match 100.0%; Score 39; DB 1; Length 17;  
Best Local Similarity 100.0%; Pred. No. 0.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 8  
Db 4 KKKKKKA 11

RESULT 3  
W03642  
ID W03642 standard; peptide; 116 AA.  
AC W03642;  
DT 17-DEC-1996 (first entry)  
DE Human cannabinoid GPR N-terminal sequence.  
KW G-protein coupled receptor; ligand binding assay; transmembrane domain;  
KW schizophrenia; dopamine; CAMP; adenosine; thrombin; adrenergic; opsin;  
KW muscarinic acetylcholine; endothelin; bombesin; endocrine; rhodopsin;  
KW odorant; cytomagalovirus; serotonergic.  
OS Homo sapiens.  
PN US5508384-A.  
PD 16-APR-1996.  
PF 10-SEP-1992; 943236.  
PR 09-SEP-1993; US-118270.  
PI (UYNV ) UNIV NEW YORK STATE.  
PI Murphy RB, Schuster DI;  
DR WPI; 96-208785/21.  
PT New dopamine receptor peptide - useful as antipsychotic agent, e.g.  
PT for treating schizophrenia  
PS Disclosure; Fig 8B(3); 184pp; English.  
CC Proteins W02657-W02720 represent a range of G-protein coupled receptor  
(GPR) proteins selected from CAMP, adenosine, muscarinic acetylcholine,  
adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,  
odorant, cytomagaloviral and other GPR proteins. The peptides  
CC W03578-W03651 represent the N-terminal fragments of the above proteins.  
CC The receptor proteins were used to design polypeptides, pref. based on  
CC the transmembrane domains, for use in G-protein coupled receptor ligand  
CC binding assays. The polypeptide fragments retain biological activity  
CC such as binding a GPR ligand or modulating GPR ligand binding to a GPR  
CC (see W02747-W02699 for examples of polypeptide fragments).  
CC The polypeptide fragments can be used in compositions for treating  
CC subjects suffering from a pathology related to a GPR abnormality e.g. a  
CC psychotic disorder such as schizophrenia.  
SQ Sequence 116 AA;

Query Match 92.3%; Score 36; DB 1; Length 116;  
Best Local Similarity 87.3%; Pred. No. 11;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 8  
Db 48 KKKKKKS 55

RESULT 4  
W24559  
ID W24559 standard; Protein; 1040 AA.  
AC W24559;  
DT 06-FEB-1998 (first entry)  
DE Presenilin-interacting protein GR24.  
KW Presenilin-interacting protein; human; Alzheimer's disease;  
OS diagnosis; therapy; transgenic animal; animal model; GR24.  
KM Homo sapiens.  
FH Key Location/Qualifiers  
FT 346..862  
FT Domain /note="presenilin-interacting domain"

FT Misc-difference 12  
FT /note="encoded by GST"  
FT Misc-difference 40  
FT /note="encoded by CCR"  
FT Misc-difference 71  
FT /note="encoded by AKC"  
FT W09727296-A1.  
PD 31-JUL-1997.  
PF 27-JAN-1997; CA0051.  
PR 02-JAN-1997; US-034590.  
PR 26-JAN-1996; US-592541.  
PR 05-JUL-1996; US-021673.  
PR 12-JUL-1996; US-021700.  
PR 08-NOV-1996; US-029895.  
PA (HSCR-) HSC RES & DEV LP.  
PA (UTOR ) UNIV TORONTO GOVERNING COUNCIL.  
PI Fraser PE, Rommens JM, St George-Hyslop PH;  
DR WPI; 97-393684/36.  
DR N-PSDB; T79967.  
PT Presenilin-interacting protein genes - used to develop products for  
PT the diagnosis, therapy and study of Alzheimer's disease and related  
PT disorders

PS Claim 1; Page 99-101; 133pp; English.  
CC GR24, a human protein with 'armadillo' repeats, has been identified  
CC as a presenilin-interacting protein (PIP). A yeast two-hybrid kit  
CC was employed to screen a human brain cDNA library for clones which  
CC interacted with presenilin transmembrane 6-7 loop domain; mutations  
CC in this loop domain are known to be causative of Alzheimer's  
CC disease (AD). 9 PIP gene sequences (see T79966-74) including GR24  
CC were identified. PIP nucleic acids, PIP proteins and peptides  
CC (especially the presenilin interacting domain), antibodies to PIPs,  
CC cells transformed with PIP nucleic acids, and transgenic animals  
CC altered with PIP nucleic acids can be used for the diagnosis,  
CC therapy and study of AD and related disorders. They can be used to  
CC identify compounds which can modulate the expression of a PIP gene  
CC or which bind to a PIP or modulate its activity.  
SQ Sequence 1040 AA;

Query Match 92.3%; Score 36; DB 1; Length 1040;  
Best Local Similarity 87.3%; Pred. No. 69;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 8  
Db 626 KKKKKKS 633

RESULT 5  
W60664  
ID W60664 standard; Protein; 756 AA.  
AC W60664;  
DT 26-OCT-1998 (first entry)  
DE Human ATRM protein.  
KW ATRM; adherens-junction linked arm protein; delta-catenin;  
KW presenilin 1; diagnosis; Alzheimer's disease; antibody; human.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT 159..202  
FT Region /note="arm repeat 1"  
FT 203..243  
FT Region /note="arm repeat 11"  
FT 412..450  
FT Region /note="arm repeat 111"  
FT 459..499  
FT Region /note="arm repeat 1v"  
FT W09825142-A1.  
PD 11-JUN-1998.  
PR 02-DEC-1997; U22093.  
PR 02-DEC-1996; US-031556.  
PA (BGHM ) BRIGHAM & WOMENS HOSPITAL.  
PI Kosik KS, Zhou J;  
DR WPI; 98-333494/29.

DR N-PSDB; V38305.  
PT Human ALARM polypeptide(s) binding to presenilin 1 - useful, e.g. to  
PT detect presenilin 1 and diagnose human diseases caused by mutant  
PT form of protein interacting with ALARM  
PS Claim 4: Fig 1A-F; 62pp; English.  
CC This novel human protein, named ALARM or delta-catenin, is able to  
CC bind to presenilin 1. Its amino acid sequence was deduced from a  
CC cDNA clone (see V38305) isolated from a human brain cDNA library.  
CC ALARM (named for adherens-junction linked arm protein) is expressed  
CC almost exclusively in brain tissue. It contains 4 copies of the  
CC arm repeat that is characteristic of catenin proteins. Overall, the  
CC ALARM and delta-catenin are 52.3% similar and 32.1% identical. The  
CC invention provides isolated ALARM polypeptide, nucleic acids,  
CC vectors, host cells and specific antibodies. Presenilin 1 protein  
CC can be detected in samples (especially human cerebrospinal fluid),  
CC by contacting with ALARM (claimed). Mutation of the presenilin 1  
CC gene is associated with familial Alzheimer's disease. ALARM can be  
CC used to diagnose human diseases caused by a mutant form of a protein  
CC interacting with ALARM, by analysing fluid samples to detect the  
CC ALARM-interacting protein (claimed). ALARM can also be used to  
CC diagnose altered levels of presenilin 1, by contacting a sample with  
CC ALARM and determining bonding (claimed). ALARM polypeptides and  
CC antibodies can be used to detect an ALARM-containing complex in  
CC biological samples, by contacting with polypeptide or antibody and  
CC determining whether the polypeptide/antibody binds to a sample  
CC component (claimed). The antibodies are also useful to detect ALARM  
CC polypeptides and to measure the effect of candidate compounds on  
CC expression or localisation of ALARM. They are useful to evaluate  
CC engineered cells prior to introduction in gene therapy, to inhibit  
CC abnormal ALARM activity or to generate anti-idiotypic antibodies,  
CC useful therapeutically.  
SQ Sequence 756 AA;

Query Match 92.3%; Score 36; DB 1; Length 756;  
Best Local Similarity 87.5%; Pred. No. 53;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KKKKKRKA 8  
| | | | | | | |  
Db 378 KKKKKRKS 385

RESULT 6  
ID P94265 standard; protein; 783 AA.  
AC P94265;  
DT 24-JUN-1990 (first entry)  
DE Sequence of APh36.1 clone.  
KW Clone APh36.1; acyl-peptide hydrolase.  
OS Rattus rattus.  
FH Key location/Qualifiers  
FT protein 1..721  
FT active-site /label=claimed protein  
FT region 580..587  
FT region /note="potential site of glycosylation"  
FT region 131.  
FT region /note="potential site of glycosylation"  
FT region 132.  
FT region /note="potential site of glycosylation"  
FT region 229.  
FT region /note="potential site of glycosylation"  
FT region 230.  
FT region /note="potential site of glycosylation"  
FT region 231.  
FT region /note="potential site of glycosylation"  
FT region 239.  
FT region /note="potential site of glycosylation"  
FT region 240.  
FT region /note="potential site of glycosylation"  
FT region 241.  
FT region /note="potential site of glycosylation"

PN EP-303997-A.  
PD 22-FEB-1989.  
PE 13-AUG-1988; 113186.  
PR 21-AUG-1987; US-087936.  
PA (GEO). The General Hospital Corp.  
PI Smith JA;  
DR WPI; 89-055444/08.  
DR N-PSDB; N91052.  
PT Recombinant acyl-peptide hydrolase -used to catalyse hydrolysis of N-acyl  
PT peptide(s) or reaction of N-acyl amino acid donor and acceptor protein.  
PS Disclosure: pp; English.  
CC Sequence contains the claimed sequence of acyl-peptide hydrolase (APH)  
CC (claim 1, page 11). APH can be used to catalyse the hydrolysis of the  
CC N-terminal acyl amino acid of an acylated polypeptide, or the reaction  
CC between a derivatised N alpha-acetyl amino acid donor and acceptor with  
CC a free alpha-NH2. It can also be used to make refractory proteins  
CC susceptible to Edman sequencing or to reduce degradation of proteins to  
CC be used therapeutically.  
SQ Sequence 783 AA;

Query Match 89.7%; Score 35; DB 1; Length 783;  
Best Local Similarity 100.0%; Pred. No. 77;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KKKKKR 7  
| | | | | | |  
Db 777 KKKKKR 783

RESULT 7  
ID P81006 standard; protein; 744 AA.  
AC P81006;  
DT 15-OCT-1990 (first entry)  
DE Alpha-2-plasmin inhibitor cDNA gene product.  
KW Alpha-2-plasmin inhibitor; APHL; plasmin; antithromorrhagic.  
OS Homo sapiens.  
PN EP-257630-A.  
PD 2-MAR-1988.  
PE 26-AUG-1987; 112381.  
PR 27-AUG-1986; JP-199056.  
PR 5-DEC-1986; JP-288875.  
PA (FARR) Hoechst Japan Ltd.  
PI Tone M, Iwaki A, Kikuno R, Hashimoto T, Okazaki H;  
DR WPI; 88-057872/09.  
DR N-PSDB; N81329.  
PT New DNA coding for human alpha-2-plasmin inhibitor like protein -  
PT used for prodn. of inhibitor for use as antithromorrhagic agent  
PT or component of tissue adhesive.  
PS Disclosure; Table 2; 31pp; English.  
CC cDNA is isolated from the HepG2 cell line.  
CC AAs marked as X are either ochre or amber stop codons.  
SQ Sequence 744 AA;

Query Match 89.7%; Score 35; DB 1; Length 744;  
Best Local Similarity 100.0%; Pred. No. 74;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KKKKKR 7  
| | | | | | |  
Db 736 KKKKKR 742

RESULT 8  
ID R08261 standard; protein; 175 AA.  
AC R08261;  
DT 18-FEB-1991 (first entry)  
DE Peptide antigenic for malarial antiparasite Antibodies.  
KW Malaria; vaccine; universal carrier molecule.  
OS Synthetic.

Key Location/Qualifiers  
FT duplication 1. .160  
FT /label- (-Asn-Ala-Ala-Gly)-3-40  
FT 161. .175  
FT duplication /label- (-Lys)-1-15 (odd number)  
FT 161. .175  
FT cross\_links /label- nth Lys has side chain identical to  
FT AAs 1. .n-1  
PN EP-398443-A.  
PD 22-NOV-1990.  
PE 16-MAY-1990; 201244.  
PR 19-MAY-1989; IT-020553.  
PR 23-MAR-1990; IT-019800.  
PA (ENIE ) ENIRICERCHE SPA.  
PI Pessi A, Bonelli F, Chiapinelli L, Bianchi E, Del Giudice G;  
DR WPI; 90-350221/47.  
PT New immunogenic cpds. for synthetic vaccines for Plasmodium  
PT malariae and for determ. of antibodies of Plasmodium malariae  
PS Claim 1; Page 15; 19pp; English.  
CC Peptides have a fractal structure, branching at each lysine residue  
CC into a side chain identical to all AAs at its N-terminal.  
CC Peptides are useful as universal carrier molecules inducing  
CC genetically unrestricted protective immunity against different  
CC pathogenic agents.  
CC Specifically, the peptides may be used as antigens in vaccination  
CC and immunoassay of Plasmodium malariae antiporozoite antibodies.  
CC See also R08260.  
SQ Sequence 175 AA;

Query Match 89.7%; Score 35; DB 1; Length 175;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKK 7  
DB 161 KKKKKK 167

RESULT 9  
R24220  
ID R24220 standard; Protein; 12 AA.  
AC R24220;  
DT 18-NOV-1992 (first entry)  
DE Fragment of tenascin-related peptide.  
KW Tenascin; related peptide; cell attachment; angiogenesis;  
KW tumour metastasis; solid matrix; prosthetic device; vascular graft;  
KW percutaneous device.  
PN W09207872-A.  
PD 14-MAY-1992.  
PE 29-OCT-1991; U08018.  
PR 29-OCT-1990; US-605920.  
PR 30-OCT-1990; US-605667.  
PA (CALB-) CALIFORNIA INST BIOLOGICAL RES.  
PI Bourdon MA;  
DR WPI; 92-183625/22.  
PT New tenascin-related peptides - modulate cell attachment to  
PT tenascin, useful in inhibition of tumour metastasis and  
PT angiogenesis  
PT Disclosure: page 8; 60pp; English.  
PS The peptide may form an N- or C-terminal fragment of the generic  
CC peptide of R24192, which is a tenascin-related peptide. This  
CC peptide mimics the ability of tenascin to promote cell attachment.  
CC The peptide and antibodies raised to it can be used to modulate cell  
CC attachment to tenascin, esp. to inhibit tumour metastasis and  
CC angiogenesis. The peptide is pref. attached to a solid matrix, eg  
CC collagen, nitrocellulose, polyester, glass, synthetic resin, long-chain  
CC polysaccharide or synthetic resin fibre. It is esp. operatively linked  
CC to a solid matrix forming a prosthetic device, percutaneous device,  
CC vascular graft, etc. For topical admin. it is formulated into a  
CC lotion, saline, gel, colloid, powder etc.  
SQ Sequence 12 AA;

Query Match 89.7%; Score 35; DB 1; Length 12;  
Best Local Similarity 100.0%; Pred. No. 2.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKK 7  
DB 1 KKKKKK 7

RESULT 10  
P71165  
ID P71165 standard; protein; 12 AA.  
AC P71165;  
DT 04-APR-1991 (first entry)  
DE Peptide which inhibits the binding of fibrinogen to platelets.  
KW Fibrinogen; platelets; cancer; cell attachment.  
PN EP-220957-A.  
PD 06-MAY-1987.  
PE 27-OCT-1986; 308335.  
PR 28-OCT-1985; US-791872.  
PA (SCRI-) SCRIPPS CLINIC RES.  
PI Zimmerman TS, Ruggeri ZM, Houghten R;  
DR WPI; 87-124477/18.  
PT New peptide(s) - useful for modulating cell adhesion and  
PT inhibiting binding of adhesive proteins to blood platelets esp.  
PT in treating cancers  
PS Claim 19; page 7; 7pp; English.  
CC The peptide inhibits the binding of fibrinogen to platelets and inhibits  
CC the aggregation of cells or platelets to each other since it modulates  
CC cell adhesion and inhibits the binding of adhesive proteins to blood  
CC platelets. The peptide has superior activity to known peptides and has  
CC similar properties. The peptide is useful for therapeutic and diagnostic  
CC purposes, esp. for inhibiting cell-cell attachment and cell growth  
CC phenomena such as cancer.  
SQ Sequence 12 AA;

Query Match 89.7%; Score 35; DB 1; Length 12;  
Best Local Similarity 100.0%; Pred. No. 2.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKK 7  
DB 1 KKKKKK 7

RESULT 11  
R11512  
ID R11512 standard; peptide; 40 AA.  
AC R11512;  
DT 21-JUN-1991 (first entry)  
DE Anti-AT1a antibody-binding peptide #1.  
KW adult T cell leukaemia associated antigen; HTLV-1; ATL.  
OS Synthetic.  
PN EP-423649-A.  
PD 24-APR-1991.  
PE 12-OCT-1990; 119624.  
PR 13-OCT-1989; JP-266983.  
PA (KORS ) KURARAY KK.  
PI Meeda Y, Shiraki H, Washitani Y, Kuroda N, Yamada K;  
DR WPI; 91-118942/17.  
PT New peptide(s) able to bind to anti-AT1a antibody - used in  
PT diagnosis and to treat HTLV-1-associated myelopathy and diseases  
PT caused by ATL.  
PS Claim 2; Page 31; 44 pp; English.  
CC This peptide is able to bind to an antibody specific against an  
CC adult T cell leukaemia associated antigen. The peptide can have  
CC from 1 to 10 Lys residues at the N-terminus. The invention also  
CC covers an adsorbent for anti-AT1a antibodies, comprising the peptide  
CC immobilised on a carrier and a reagent containing the peptide, for  
CC measuring the antibody. The reagent and the adsorbent can be used

CC to treat an adult T-cell leukaemia virus-infectious disease  
CC See also R1513 and R1514.  
SQ Sequence 40 AA;

```
Query Match      89.7%; Score 35; DB 1; Length 40;
Best Local Similarity . 100.0%; Pred.No. 6.5;
Matches    7; Conservative    0; Mismatches    0; Indels    0; Gaps    0;
```

RESULT 12  
R11513  
ID R11513 standard: Protein: 29 AA.  
AC R11513;  
DT 21-JUN-1991 (first entry)  
DE Anti-ATL/A antibody-binding peptide #2.  
KW adult T cell leukemia associated antigen; HTLV-1; ATL.  
OS Synthetic.  
PN EP-423649-A.  
PD 24-APR-1991.  
PE 12-OCT-1990; 119624.  
PR 13-OCT-1989, JP-266983.  
PA (KURAS ) KURARAY KK.  
PI Maeda Y, Shiraki H, Washitani Y, Kuroda N, Yamada K;  
PI Oka K, Namba T.  
DR WPI: 91-118942/17.  
PT New peptide(s) able to bind to anti-ATLA antibody - used in  
PT diagnosis and to treat HTLV-1-associated myelopathy and diseases  
PT caused by ATL.  
PS Claim 3: Page 31: 44 pp: English.  
CC This peptide is able to bind to an antibody specific against an  
CC adult T cell leukemia associated antigen. The peptide can have  
CC from 1 to 10 lys residues at the N-terminus. The invention also  
CC covers an adsorbent for anti-ATLA antibodies, comprising the peptide  
CC immobilised on a carrier and a reagent containing the peptide, for  
CC measuring the antibody. The reagent and the adsorbent can be used  
CC to treat an adult T-cell leukemia virus-infectious disease.  
CC See also R11512 and R11514.  
Q Sequence 29 AA;

Query Match	89.7%	Score 35;	DB 1;	Length 29;
Best Local Similarity	100.0%	Pred. No. 5;		
Matches 7;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	KKKKKKK	7
Db	1	KKKKKKK	7

RESULT	13
ID	R11514
AC	R11514; standard: Protein; 34 AA.
DT	21-JUN-1991 (first entry)
DE	Anti-ATLA antibody-binding peptide #3.
KW	adult T cell leukaemia associated antigen; HTLV-1; ATL.V.
OS	Synthetic.
PN	EP-423648-A.
PD	24-APR-1991.
PF	12-OCT-1990; 119624.
PR	13-OCT-1989; JP-266983.
PA	(KURS ) KURARAY KK.
PI	Meda Y, Shiraki H, Washitani Y, Kuroda N, Yamada K,
PI	Oka K, Namba T:
DR	WPI, 91-118942/17.
PT	New peptide(s) able to bind to anti-ATLA antibody - used in
PT	diagnosis and to treat HTLV-1-associated myelopathy and diseases
PT	caused by ATL.V

PS Claim 4 page 31: 44 pp: English.  
CC This peptide is able to bind to an antibody specific against an  
CC adult T cell leukemia associated antigen. The peptide can have  
CC from 1 to 10 lys residues at the N-terminus. The invention also  
CC covers an adsorbent for anti-ATLA antibodies, comprising the peptide  
CC immobilised on a carrier and a reagent containing the peptide, for  
CC measuring the antibody. The reagent and the adsorbent can be used  
CC to treat an adult T-cell leukemia virus-infectious disease.  
CC See also R11512 and R11513.  
CC Sequence 34 AA;

		89.7%;	Score 35;	DB 1;	Length 34;
Query Match					
Best Local Similarity		100.0%;	Pred. No. 5.7;		
Matches	7;	Conservative	0;	Mismatches	0; Indels
					Gaps 0.
QY	1 KKKKKK	7			
Dd	1 KKKKKK	7			

RESULT	14
P61030	
ID	p61030 standard; Protein: 898 AA.
AC	p61030;
DT	13-AUG-1991 (first entry)
DE	Entire coded sequence from plasmid pAU157 insert encoding rat-liver
DE	P-450MC cytochrome.
KW	3-methyl-choleanthrene; MC; ds.
OS	Rattus sp.
PN	J61032284-A.
PD	14-MAR-1986.
PF	15-AUG-1984; 169447.
PR	15-AUG-1984; JP-169447.
PR	24-AUG-1984; JP-175159.
PA	(AGEN ) AGENCY OF IND SCI TECH.
DR	WPI; 86-109962/17.
DR	N-PSDS; N60909.
DT	New plasmid used in coding rat-liver cytochrome P-450 gene.
PS	Disclosure; Fig 1; 13pp; Japanese.
CC	The gene product may be produced in commercial quantities from a
CC	transformed expression system. The rat-liver cytochrome is readily
CC	separated, and is induced by 3-methyl-choleanthrene.
CC	As marked as x correspond to stop codons in the insert.
Q	Sequence 898 AA;

Query Match	89.7%	Score 35:	DB 1;	Length 898;
Best Local Similarity	100.0%	Pred. NO.	87;	
Matches	7;	Conservative	0;	Mismatches 0;
				Gaps 0;

QY	1	KKKKKKK	7
Db	873	KKKKKKK	879

RESULT	15
PT	cytochrome P-450MC gene in Escherichia coli.
DR	N-FSDB; N06064.
PA	(AGEN) AGENCY OF IND SCI TECH.
PR	16-JUN-1984; JP-122952.
PF	16-JUN-1984; 122952.
PD	11-JAN-1986.
PN	J61005783-A.
OS	Rattus rattus.
KW	Organic oxide removal; tac-promoter.
DE	Translation of plasmid pAU157 encoding rat liver cytochrome P-450MC.
AC	P61056; 08-JUL-1991 (first entry)
ID	P61056 standard; Protein; 899 AA.
PT	P61056

PS Disclosure; Fig 1; 9pp; Japanese.  
 CC The plasmid may be expressed from a transformed E.coli host, under  
 CC the control of a strong upstream tac-promoter. The product  
 CC may be useful in the removal of oxides from organic compounds in  
 CC industrial waste by an oxidation reaction.  
 CC In this translation, all stop codons are given as unknown (x).  
 SQ Sequence 899 AA;

Query Match 89.7%; Score 35; DB 1; Length 899;  
 Best Local Similarity 100.0%; Pred. No. 87;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKK 7  
 |||||  
 Db 873 KKKKKK 879

Search completed: July 18, 1999, 00:42:53  
 Job time: 8586 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 06:07:55 ; Search time 39.54 Seconds  
(without alignments)  
1.997 Million cell updates/sec

Title: US-09-142-043-2  
Perfect score: 39  
Sequence: 1 KKKKKKA 8

Scoring table: BLOSUM62

Searched: 106577 seqs, 9868381 residues

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/PCTUS9.COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/Backfill1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	35	89.7	11	1	US-07-694-983-15	Sequence 15, Appl
2	35	89.7	473	1	US-08-103-739B-2	Sequence 2, Appl
3	35	89.7	29	1	US-08-054-363-17	Sequence 17, Appl
4	35	89.7	36	1	US-07-948-357-4	Sequence 4, Appl
5	35	89.7	36	1	US-07-948-357-12	Sequence 12, Appl
6	35	89.7	36	1	US-07-948-357-13	Sequence 13, Appl
7	35	89.7	1235	1	US-08-118-101A-2	Sequence 2, Appl
8	35	89.7	29	1	US-08-088-658-22	Sequence 22, Appl
9	35	89.7	10	1	US-08-097-830E-1	Sequence 1, Appl
10	35	89.7	10	1	US-08-097-830E-2	Sequence 2, Appl
11	35	89.7	434	1	US-08-097-830E-3	Sequence 3, Appl
12	35	89.7	25	1	US-08-240-514-56	Sequence 56, Appl
13	35	89.7	1497	1	US-08-623-679-7	Sequence 7, Appl
14	35	89.7	1533	1	US-08-623-679-9	Sequence 9, Appl
15	35	89.7	398	1	US-08-074-431-2	Sequence 2, Appl
16	35	89.7	7	1	US-08-704-170-11	Sequence 11, Appl
17	35	89.7	29	2	US-08-595-387-17	Sequence 17, Appl
18	35	89.7	182	2	US-08-226-264-28	Sequence 28, Appl
19	35	89.7	25	2	US-08-612-302A-56	Sequence 56, Appl
20	35	89.7	473	2	US-08-474-404-2	Sequence 2, Appl
21	35	89.7	20	2	US-08-769-211-1	Sequence 1, Appl
22	35	89.7	20	2	US-08-769-211-4	Sequence 4, Appl
23	35	89.7	10	2	US-08-456-112B-1	Sequence 1, Appl
24	35	89.7	30	2	US-08-456-112B-2	Sequence 2, Appl
25	35	89.7	434	2	US-08-456-112B-3	Sequence 3, Appl
26	35	89.7	473	2	US-08-485-845-2	Sequence 2, Appl
27	35	89.7	473	2	US-08-482-714-2	Sequence 2, Appl
28	35	89.7	394	2	US-08-951-148-5	Sequence 5, Appl
29	35	89.7	394	2	US-08-951-148-10	Sequence 10, Appl
30	35	89.7	398	2	US-08-902-655A-2	Sequence 2, Appl
31	35	89.7	478	2	US-09-040-799-3	Sequence 3, Appl
32	35	89.7	1160	3	PCT-US82-05401-4	Sequence 4, Appl
33	35	89.7	7	3	PCT-US94-02631-11	Sequence 11, Appl
34	35	89.7	24	3	PCT-US95-06077-23	Sequence 23, Appl
35	35	89.7	191	3	PCT-US95-06077-24	Sequence 24, Appl
36	35	89.7	191	3	PCT-US95-07171-3	Sequence 3, Appl
37	35	89.7	1087	1	US-08-264-002-5	Sequence 5, Appl
38	32	82.1	723	1	US-07-814-964-11	Sequence 11, Appl
39	32	82.1	723	1	US-08-258-442-11	Sequence 11, Appl

40	32	82.1	723	1	US-08-328-809-6	Sequence 6, Appl
41	32	82.1	8	2	US-08-701-124-1	Sequence 1, Appl
42	32	82.1	11	2	US-08-701-124-19	Sequence 19, Appl
43	32	82.1	21	2	US-08-701-124-27	Sequence 27, Appl
44	32	82.1	15	2	US-08-701-124-43	Sequence 43, Appl
45	32	82.1	723	3	PCT-US92-11107-11	Sequence 11, Appl

## ALIGNMENTS

```

RESULT 1
US-07-694-983-15
; Sequence 15, Application US/07694983
; Patent No. 5432260
; GENERAL INFORMATION:
; APPLICANT: Stahl, Philip D.
; TITLE OF INVENTION: HIGH AFFINITY MANNOSE RECEPTOR
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; STREET: 345 Middlefield Road, Suite 200
; CITY: Menlo Park
; STATE: California
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/694,983
; FILING DATE: 19910503
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 9500-0039.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-327-7250
; TELEFAX: 415-327-2951
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1
; OTHER INFORMATION: /label= Ac-
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 11
; OTHER INFORMATION: /label= -NH2
; US-07-694-983-15

Query Match      89.7%  Score 35; DB 1; Length 11;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKK 7
   |||||
Db 2 KKKKKK 8

RESULT 2
US-08-103-739B-2

```

Sequence 2, Application US/08103739B  
Patent No. 5478369  
GENERAL INFORMATION:  
APPLICANT: ALBERTSEN, Marc C.  
APPLICANT: BEACH, Larry R.  
APPLICANT: HOWARD, John A.  
APPLICANT: HUFFMAN, Gary A.  
TITLE OF INVENTION: DNA Sequences Mediating Male Fertility  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pioneer Hi-Bred International, Inc.  
STREET: 700 Capital Square, 400 Locust Street  
CITY: Des Moines  
STATE: Iowa  
COUNTRY: U.S.  
ZIP: 50309  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/103,739B  
FILING DATE: 02-AUG-1993  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/537,183  
FILING DATE: 12-JUN-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Sweeney, Patricia A.  
REGISTRATION NUMBER: 32,733  
REFERENCE/DOCKET NUMBER: 0125R2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (515) 248 4897  
TELEFAX: (515) 248 4844  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 473 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-103-739B-2

Query Match 89.7%; Score 35; DB 1; Length 473;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKK 7  
111111  
Db 465 KKKKKK 471

RESULT 3  
US-08-054-363-17  
Sequence 17, Application US/08054363  
Patent No. 5539082  
GENERAL INFORMATION:  
APPLICANT: Nielsen, Peter E.  
APPLICANT: Buchardt, Ole  
APPLICANT: Egholm, Michael  
APPLICANT: Berg, Rolf H.  
TITLE OF INVENTION: No. 5539082a1 Peptide Nucleic Acids  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock, Mashburn, Kurtz, Mackiewicz, and No. 5539082a1s  
STREET: One Liberty Place - 46th Floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/054,363  
FILING DATE: 19930426  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Lucci, Joseph  
REGISTRATION NUMBER: 33,307  
REFERENCE/DOCKET NUMBER: IS1S1017  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-568-3100  
TELEFAX: 215-568-3439  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 29 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1  
OTHER INFORMATION: /label- Modified-site  
OTHER INFORMATION: /note- "Cytosine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 2  
OTHER INFORMATION: /label- Modified-site  
OTHER INFORMATION: /note- "Cytosine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 3  
OTHER INFORMATION: /label- Modified-site  
OTHER INFORMATION: /note- "Cytosine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 4  
OTHER INFORMATION: /label- Modified-site  
OTHER INFORMATION: /note- "Cytosine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 5  
OTHER INFORMATION: /label- Modified-site  
OTHER INFORMATION: /note- "Cytosine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 6  
OTHER INFORMATION: /label- Modified-site  
OTHER INFORMATION: /note- "Adenine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 9 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 7  
OTHER INFORMATION: /label- Modified-site  
OTHER INFORMATION: /note- "Cytosine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 8  
OTHER INFORMATION: /label- Modified-site



OTHER INFORMATION: /note- "Cytosine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 9  
OTHER INFORMATION: /label- Modified-site  
OTHER INFORMATION: /note- "Adenine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 9 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 10  
OTHER INFORMATION: /label- Modified-site  
OTHER INFORMATION: /note- "Cytosine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 11  
OTHER INFORMATION: /label- Modified-site  
OTHER INFORMATION: /note- "Thymine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 12  
OTHER INFORMATION: /label- Modified-site  
OTHER INFORMATION: /note- "Thymine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 13  
OTHER INFORMATION: /label- Modified-site  
OTHER INFORMATION: /note- "Cytosine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 14  
OTHER INFORMATION: /label- Modified-site  
OTHER INFORMATION: /note- "Cytosine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 15  
OTHER INFORMATION: /label- Modified-site  
OTHER INFORMATION: /note- "Cytosine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 16  
OTHER INFORMATION: /label- Modified-site  
OTHER INFORMATION: /note- "Cytosine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 17  
OTHER INFORMATION: /label- Modified-site  
OTHER INFORMATION: /note- "Thymine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 18  
OTHER INFORMATION: /label- Modified-site  
OTHER INFORMATION: /note- "Cytosine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."

FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 19  
OTHER INFORMATION: /label- Modified-site  
OTHER INFORMATION: /note- "Thymine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 20  
OTHER INFORMATION: /label- Modified-site  
OTHER INFORMATION: /note- "Cytosine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
US-08-054-363-17  
Query Match 89.7%; Score 35; DB 1; Length 29;  
Best Local Similarity 100.0%; Pred. No. 4.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KKKKKK 7  
Db 21 KKKKKK 27  
RESULT 4  
US-07-948-357-4  
Sequence 4, Application US/07948357  
Patent No. 5347932  
GENERAL INFORMATION:  
APPLICANT: Curjel, David J.  
APPLICANT: Birnstiel, Max L.  
APPLICANT: Cotton, Matthew  
APPLICANT: Wagner, Ernst  
APPLICANT: Zatloukal, Kurt  
APPLICANT: Plank, Christian  
APPLICANT: Oberhauser, Berndt  
APPLICANT: Schmidt, Walter G.M.  
TITLE OF INVENTION: Composition for Introducing Nucleic Acid  
TITLE OF INVENTION: Complexes Into Higher Eucaryotic Cells  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1225 Connecticut Avenue  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/948,357  
FILING DATE: 19920923  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Esmond, Robert W.  
REGISTRATION NUMBER: 32, 893  
REFERENCE/DOCKET NUMBER: 0652, 0940004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 466-0800  
TELEFAX: (202) 833-8716  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 36 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: both  
MOLECULE TYPE: peptide  
US-07-948-357-4

Query Match 89.7%; Score 35; DB 1; Length 36;  
Best Local Similarity 100.0%; Pred. No. 4.8;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKK 7  
111111  
Db 25 KKKKKK 31

RESULT 5  
US-07-948-357-12  
Sequence 12, Application US/07948357  
Patent No. 5547932

GENERAL INFORMATION:  
APPLICANT: Curtiel, David T.  
APPLICANT: Birstiel, Max L.  
APPLICANT: Cotten, Matthew  
APPLICANT: Wagner, Ernst  
APPLICANT: Zatloukal, Kurt  
APPLICANT: Plank, Christian  
APPLICANT: Oberhauser, Berndt  
TITLE OF INVENTION: Composition for Introducing Nucleic Acid  
TITLE OF INVENTION: Complexes Into Higher Eucaryotic Cells  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1225 Connecticut Avenue  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk.  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/948,357  
FILING DATE: 19920923  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Esmond, Robert W.  
REGISTRATION NUMBER: 32,893  
REFERENCE/DOCKET NUMBER: 0652,0940004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 466-0800  
TELEFAX: (202) 833-8716  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 36 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: both  
MOLECULE TYPE: peptide  
US-07-948-357-12

Query Match 89.7%; Score 35; DB 1; Length 36;  
Best Local Similarity 100.0%; Pred. No. 4.8;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKK 7  
111111  
Db 27 KKKKKK 33

RESULT 6  
US-07-948-357-13  
Sequence 13, Application US/07948357  
Patent No. 5547932

GENERAL INFORMATION:

APPLICANT: Curtiel, David T.  
APPLICANT: Birstiel, Max L.  
APPLICANT: Cotten, Matthew  
APPLICANT: Wagner, Ernst  
APPLICANT: Zatloukal, Kurt  
APPLICANT: Plank, Christian  
APPLICANT: Oberhauser, Berndt  
TITLE OF INVENTION: Composition for Introducing Nucleic Acid  
TITLE OF INVENTION: Complexes Into Higher Eucaryotic Cells  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1225 Connecticut Avenue  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/948,357  
FILING DATE: 19920923  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Esmond, Robert W.  
REGISTRATION NUMBER: 32,893  
REFERENCE/DOCKET NUMBER: 0652,0940004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 466-0800  
TELEFAX: (202) 833-8716  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 36 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: both  
MOLECULE TYPE: peptide  
US-07-948-357-13

Query Match 89.7%; Score 35; DB 1; Length 36;  
Best Local Similarity 100.0%; Pred. No. 4.8;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKK 7  
111111  
Db 27 KKKKKK 33

RESULT 7  
US-08-118-101A-2  
Sequence 2, Application US/08118101A  
Patent No. 5620892

GENERAL INFORMATION:  
APPLICANT: Kutz, Stephen E.  
APPLICANT: Knickerbocker, Aron M.  
APPLICANT: McCullough, John R.  
TITLE OF INVENTION: A STRAIN OF SACCHAROMYCES CEREVISIAE  
TITLE OF INVENTION: EXPRESSING THE GENE ENCODING POTASSIUM TRANSPORTER MINK  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burton Rodney  
STREET: P.O. Box 4000  
CITY: Princeton  
STATE: New Jersey  
COUNTRY: U.S.A.  
ZIP: 08543-4000  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/118,101A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Gaul, Timothy J.  
REGISTRATION NUMBER: 33,111  
REFERENCE/DOCKET NUMBER: DC27  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (609) 252-5901  
TELEFAX: (609) 252-4526  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1235 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-118-101A-2

Query Match 89.7%; Score 35; DB 1; Length 1235;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKK 7  
Db 1052 KKKKKK 1058

RESULT 8  
US-08-088-658-22  
Sequence 22, Application US/0808658  
Patent No. 5641625  
GENERAL INFORMATION:  
APPLICANT: Ecker, David J.  
APPLICANT: Buchardt, Ole  
APPLICANT: Egholm, Michael  
APPLICANT: Nielsen, Peter E.  
APPLICANT: Berg, Rolf H.  
APPLICANT: M Ilegard, Niels E.  
TITLE OF INVENTION: HIGH ORDER STRUCTURE AND BINDING OF PEPTIDE  
TITLE OF INVENTION: NUCLEIC ACIDS  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz and No. 5641625RIS  
STREET: One Liberty Place - 46th Floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: U.S.A.  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/088,658  
FILING DATE: 19930702  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/054,363  
FILING DATE: 26-APRIL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Luccl, Joseph  
REGISTRATION NUMBER: 33,307  
REFERENCE/DOCKET NUMBER: ISIS-1052  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-568-3100  
TELEFAX: 215-568-3439

INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 29  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1  
OTHER INFORMATION: /label= MODIFIED-SITE  
OTHER INFORMATION: /note= "Heterocyclic base is attached to N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 9 of the base."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 2  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Heterocyclic base is attached to N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 9 of the base."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 3  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Heterocyclic base is attached to N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 9 of the base."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 4  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Heterocyclic base is attached to N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 9 of the base."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 5  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Heterocyclic base is attached to N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 9 of the base."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 6  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Heterocyclic base is attached to N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the base."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 7  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Heterocyclic base is attached to N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 9 of the base."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 8  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Heterocyclic base is attached to N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 9 of the base."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 9  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Heterocyclic base is attached to N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the base."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 10

OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Heterocyclic base is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 9 of the base."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 11  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Heterocyclic base is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the base."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 12  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Heterocyclic base is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the base."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 13  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Heterocyclic base is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 9 of the base."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 14  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Heterocyclic base is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 9 of the base."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 15  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Heterocyclic base is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 9 of the base."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 16  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Heterocyclic base is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 9 of the base."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 17  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Heterocyclic base is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the base."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 18  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Heterocyclic base is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 9 of the base."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 19  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Heterocyclic base is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the base."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 20  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Heterocyclic base is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the base."

OTHER INFORMATION: group at position 9 of the base."  
US-08-088-658-22  
Query Match 89.7%; Score 35; DB 1; Length 29;  
Best Local Similarity 100.0%; Pred. No. 4.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KKKKKK 7  
111111  
DB 21 KKKKKK 27  
RESULT 9  
US-08-097-830E-1  
Sequence 1, Application US/08097830E  
Patent No. 5652211  
GENERAL INFORMATION:  
APPLICANT: Porto, Massimo  
TITLE OF INVENTION: Peptides For Neutralizing The  
TITLE OF INVENTION: Toxicity of Lipid A  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Hedman, Gibson & Costigan, P.C.  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: DOS  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/097, 830E  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Costigan, James V.  
REGISTRATION NUMBER: 25,669  
REFERENCE/DOCKET NUMBER: 576-003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 302-8989  
TELEFAX: (212) 302-8998  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-097-830E-1  
Query Match 89.7%; Score 35; DB 1; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.7;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KKKKKK 7  
111111  
DB 1 KKKKKK 7  
RESULT 10  
US-08-097-830E-2  
Sequence 2, Application US/08097830E  
Patent No. 5652211  
GENERAL INFORMATION:  
APPLICANT: Porto, Massimo  
TITLE OF INVENTION: Peptides For Neutralizing The  
TITLE OF INVENTION: Toxicity of Lipid A

NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Hedman, Gibson & Costigan, P.C.  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: diskette, 3.50 inch, 1.44 Mb storage  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: DOS  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/097,830E  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Costigan, James V.  
REGISTRATION NUMBER: 25,669  
REFERENCE/DOCKET NUMBER: 576-003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 302-8989  
TELEFAX: (212) 302-8998  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-097-830E-2

Query Match 89.7%; Score 35; DB 1; Length 30;  
Best Local Similarity 100.0%; Pred. No. 4.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKK 7  
Db 1 KKKKKK 7

RESULT 11  
US-08-097-830E-3  
Sequence 3, Application US/08097830E  
Patent No. 5652211  
GENERAL INFORMATION:  
APPLICANT: Porro, Massimo  
TITLE OF INVENTION: Peptides For Neutralizing The  
TITLE OF INVENTION: Toxicity of Lipid A  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Hedman, Gibson & Costigan, P.C.  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: diskette, 3.50 inch, 1.44 Mb storage  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: DOS  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/097,830E  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:

NAME: Costigan, James V.  
REGISTRATION NUMBER: 25,669  
REFERENCE/DOCKET NUMBER: 576-003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 302-8989  
TELEFAX: (212) 302-8998  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 434 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-097-830E-3

Query Match 89.7%; Score 35; DB 1; Length 434;  
Best Local Similarity 100.0%; Pred. No. 3.7;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKK 7  
Db 1 KKKKKK 7

RESULT 12  
US-08-240-514-56  
Sequence 56, Application US/08240514  
Patent No. 5670347  
GENERAL INFORMATION:  
APPLICANT: GOPAL, T. Venkat  
TITLE OF INVENTION: PEPTIDE-MEDIATED GENE TRANSFER  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/240,514  
FILING DATE: 11-MAY-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 73521/102/CLIN  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 56:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-240-514-56

Query Match 89.7%; Score 35; DB 1; Length 25;  
Best Local Similarity 100.0%; Pred. No. 3.6;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKK 7  
Db 14 KKKKKK 20

RESULT 13  
US-08-623-679-7  
Sequence 7, Application US/08623679  
Patent No. 5674739  
GENERAL INFORMATION:  
APPLICANT: Shyjan, Andrew W.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
DIAGNOSIS, PREVENTION AND TREATMENT OF TUMOR  
TITLE OF INVENTION: PROGRESSION  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/623,679  
FILING DATE: 29-MAR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/412,431  
FILING DATE: 29-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Fasse, J. Peter  
REGISTRATION NUMBER: 32,983  
REFERENCE/DOCKET NUMBER: 07334/004001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1497 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-623-679-7

Query Match 89.7%; Score 35; DB 1; Length 1497;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKK 7  
Db 515 KKKKKK 521

RESULT 14  
US-08-623-679-9  
Sequence 9, Application US/08623679  
Patent No. 5674739  
GENERAL INFORMATION:  
APPLICANT: Shyjan, Andrew W.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
DIAGNOSIS, PREVENTION AND TREATMENT OF TUMOR  
TITLE OF INVENTION: PROGRESSION  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/623,679  
FILING DATE: 29-MAR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/412,431  
FILING DATE: 29-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Fasse, J. Peter  
REGISTRATION NUMBER: 32,983  
REFERENCE/DOCKET NUMBER: 07334/004001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1533 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-623-679-9

Query Match 89.7%; Score 35; DB 1; Length 1533;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKK 7  
Db 515 KKKKKK 521

RESULT 15  
US-08-507-431-2  
Sequence 2, Application US/08507431  
Patent No. 5693518  
GENERAL INFORMATION:  
APPLICANT: Kotof, Lene V.  
APPLICANT: Kauppinen, Markus S.  
APPLICANT: Christgau, Stephan  
APPLICANT: Heldt-Hansen, Hans P.  
APPLICANT: Dalboge, Henrik  
APPLICANT: Andersen, Lene N.  
APPLICANT: Sl, Joan Q.  
APPLICANT: Jacobson, Tina  
APPLICANT: Munk, Niels  
APPLICANT: Mullertz, Anette  
TITLE OF INVENTION: ENZYMES WITH XYLANASE ACTIVITY FROM  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 56935180 No. 5693518disk of No. 5693518th America, Inc.  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/507,431  
FILING DATE: 15-FEB-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/002,800  
FILING DATE: 25-AUG-1996

ATTORNEY/AGENT INFORMATION:  
 NAME: Harrington, James J.  
 REGISTRATION NUMBER: 38,711  
 REFERENCE/DOCKET NUMBER: 3954.204-US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212-867-0123  
 TELEFAX: 212-878-9655  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 398 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-507-431-2

Query Match 89.7%; Score 35; DB 1; Length 398;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKK 7  
 |||||  
 Db 389 KKKKKK 395

Search completed: July 18, 1999, 06:07:55  
 Job time: 297 sec





GenCore version 4.5  
Copyright (c) 1993 - 1998 Comugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 06:07:02 ; Search time 44.47 Seconds  
(without alignments)  
6.739 Million cell updates/sec

Title: US-09-142-043-2

Perfect score: 39

Sequence: 1 KKKKKKA 8

Scoring table: BLOSUM62

Searched: 116738 seqs, 37460341 residues

Database : PIR\_58:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	39	100.0	437	2	I51238	translation elonga
2	39	100.0	1647	2	S45252	SNF2beta protein -
3	36	92.3	370	1	S68187	choline-phosphate
4	36	92.3	1085	2	S62516	hypothetical prote
5	36	92.3	295	2	S61039	hypothetical prote
6	36	92.3	142	2	S54481	hypothetical prote
7	36	92.3	590	2	S63193	hypothetical prote
8	36	92.3	730	2	S54625	probable membrane
9	36	92.3	147	2	JQ1320	high mobility grou
10	36	92.3	886	2	I38182	transcription term
11	36	92.3	1320	2	JC5630	TCOP1 protein - mo
12	35	89.7	478	1	DEHUM2	methionyl aminopep
13	35	89.7	225	1	A25910	small nuclear ribo
14	35	89.7	1235	1	PMBYH	potassium transpor
15	35	89.7	480	2	A46702	methionyl aminopep
16	35	89.7	610	2	S71758	DEAD box protein M
17	35	89.7	482	2	S75847	signal recognition
18	35	89.7	340	2	S54181	p38-264 protein -
19	35	89.7	333	2	JC3229	translation initia
20	35	89.7	333	2	A31226	translational initia
21	35	89.7	333	2	S13147	protein synthesis
22	35	89.7	91	2	C33172	C-ORF-F protein -
23	35	89.7	1424	2	S11480	hypothetical prote
24	35	89.7	224	2	C64240	mobilitaton prote
25	35	89.7	469	2	E64319	carbon-monoxide de
26	35	89.7	506	2	S34497	glycoprotein A - P
27	35	89.7	1051	2	JC4091	hypothetical prote
28	35	89.7	246	2	S49770	hypothetical prote
29	35	89.7	139	2	S30850	hypothetical prote
30	35	89.7	144	2	S46716	hypothetical prote
31	35	89.7	167	2	S38112	hypothetical prote
32	35	89.7	130	2	S55141	hypothetical prote
33	35	89.7	445	2	S67137	hypothetical prote
34	35	89.7	1435	2	S69632	regulatory protein
35	35	89.7	319	2	B44863	RA5 antigen - Plas
36	35	89.7	418	2	S41044	chromosomal protei
37	35	89.7	1257	2	I58383	retinoblastoma bin
38	35	89.7	920	2	I58390	retinoblastoma bin
39	35	89.7	866	2	I78883	retinoblastoma bin

40	35	89.7	754	2	B42997	retinoblastoma-ass
41	35	89.7	784	2	A33633	transcription repr
42	35	89.7	671	2	A40692	signal recognition
43	35	89.7	215	2	I52523	nucleoporin p62 ho
44	35	89.7	357	2	E70333	hypothetical prote
45	35	89.7	404	2	B71224	hypothetical prote

## ALIGNMENTS

RESULT 1

I51238

translation elongation factor EF-1 gamma - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C>Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 07-Aug-1998

C:Accession: I51238

R:Morales, J.; Bassez, T.; Cormier, P.; Mulner-Lorillon, O.; Belle, R.; Osborne, H.B.

Dev. Genet. 14, 440-448, 1993

A:Title: Expression of elongation factor 1 alpha (EF-1 alpha) and 1 beta gamma (EF-1

A:Reference number: I51237; MUID:94155465

A:Accession: I51238

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Residues: 1-437 <MOR>

A:Cross-references: GB:S69726; NID:9545506; PID:9545507

C:Superfamily: translation elongation factor eEF-1 gamma chain

Query Match 100.0%; Score 39; DB 2; Length 437;

Best Local Similarity 100.0%; Pred. No. 11;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKKA 8

DB 244 KKKKKKA 251

RESULT 2

S45252

SNF2beta protein - human

C:Species: Homo sapiens (man)

C>Date: 10-Dec-1994 #sequence\_revision 17-Nov-1995 #text\_change 12-Sep-1997

C:Accession: S45252

R:Chiba, H.; Muramatsu, M.; Nomoto, A.; Kato, H.

Nucleic Acids Res. 22, 1815-1820, 1994

A:Title: Two human homologues of Saccharomyces cerevisiae SWI2/SNF2 and Drosophila br

A:Reference number: S45251

A:Accession: S45252

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1647 <CHI>

C:Superfamily: unassigned bromodomain proteins; bromodomain homology

F:1485-1540/Domain: bromodomain homology <BRO>

Query Match 100.0%; Score 39; DB 2; Length 1647;

Best Local Similarity 100.0%; Pred. No. 34;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKKA 8

DB 582 KKKKKKA 589

RESULT 3

S68187

choline-phosphate cytidylyltransferase (EC 2.7.7.15) - Plasmodium falciparum

N:Alternate names: CTP:choline-phosphate cytidylyltransferase; phosphorlycholine tran

C:Species: Plasmodium falciparum

C>Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 26-Feb-1998

C:Accession: S68187

R:Yeo, H.J.; Sri Widadana, J.; Mercereau-Pujjalton, O.; Vial, H.J.

Eur. J. Biochem. 233, 62-72, 1995  
A:Title: Molecular cloning of CTP:phosphocholine cytidylyltransferase from *Plasmodium falciparum*  
A:Reference number: S68187; PMID:96061933  
A:Accession: S68187  
A:Molecule type: DNA  
A:Residues: 1-370 <YEO>  
A:Cross-references: EMBL:X84041; NID:g1054826; PID:g1054827  
C:Genetics:  
A:Gene: CTP  
A:Introns: #status absent  
C:Function:  
A:Description: catalyzes the formation of cytidinediphosphate choline and pyrophosphate  
A:Pathway: phospholipid biosynthesis  
C:Superfamily: choline-phosphate cytidylyltransferase  
C:Keywords: membrane-associated protein; nucleotidylyltransferase; phospholipid biosynthesis  
F:274-294/Domain: amphipathic helix #status predicted <AH1>  
F:308-324/Domain: amphipathic helix #status predicted <AH2>

Query Match 92.3%; Score 36; DB 1; Length 370;  
Best Local Similarity 87.5%; Pred. No. 28;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKKA 8  
|||||||:  
Db 193 KKKKKKS 200

RESULT 4  
S62516  
Hypothetical protein loc (cosmid c2966) - fission yeast (*Schizosaccharomyces pombe*)  
C:Species: *Schizosaccharomyces pombe*  
C:Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 24-Sep-1998  
C:Accession: S62516  
R:Jones, L.; Murphy, L.; McNeill, A.; Simpson, I.; Harris, D.  
Submitted to the EMBL Data Library, October 1995  
A:Reference number: S62507  
A:Accession: S62516  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-1085 <JON>  
A:Cross-references: EMBL:Z66525; NID:g1044926; PID:g1044936  
C:Genetics:  
A:Map position: 1

Query Match 92.3%; Score 36; DB 2; Length 1085;  
Best Local Similarity 87.5%; Pred. No. 69;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKKA 8  
|||||||:  
Db 186 KKKKKKS 193

RESULT 5  
S61039  
Hypothetical protein YDL73w - yeast (*Saccharomyces cerevisiae*)  
N:Alternate names: hypothetical protein D1438  
C:Species: *Saccharomyces cerevisiae*  
C:Date: 15-Feb-1996 #sequence\_revision 01-Mar-1996 #text\_change 12-Dec-1997  
C:Accession: S61039; S67725  
R:Pohl, T.M.  
Submitted to the EMBL Data Library, November 1995  
A:Reference number: S61010  
A:Accession: S61039  
A:Molecule type: DNA  
A:Residues: 1-295 <POH>  
A:Cross-references: EMBL:Z67750; NID:g1061256; PID:g1061265  
R:Pohl, T.M.  
Submitted to the Protein Sequence Database, July 1996  
A:Reference number: S67708  
A:Accession: S67725

A:Molecule type: DNA  
A:Residues: 1-295 <POH>  
A:Cross-references: EMBL:Z74221; NID:g1431275; PID:e253081; PID:g1431276; MIPS:YDL173  
A:Experimental source: strain S288C  
C:Genetics:  
A:Map position: 4L

Query Match 92.3%; Score 36; DB 2; Length 295;  
Best Local Similarity 87.5%; Pred. No. 24;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKKA 8  
|||||||:  
Db 277 KKKKKKS 284

RESULT 6  
S54481  
Hypothetical protein YMR269w - yeast (*Saccharomyces cerevisiae*)  
N:Alternate names: hypothetical protein YMR156.11  
C:Species: *Saccharomyces cerevisiae*  
C:Date: 08-Jul-1995 #sequence\_revision 19-Oct-1995 #text\_change 21-Nov-1997  
C:Accession: S54481  
R:Lye, G.; Churcher, C.M.  
Submitted to the EMBL Data Library, May 1995  
A:Reference number: S54014  
A:Accession: S54481  
A:Molecule type: DNA  
A:Residues: 1-142 <LYE>  
A:Cross-references: EMBL:Z49260; NID:g809081; PID:g809092; MIPS:YMR269w  
A:Experimental source: strain AB972  
C:Genetics:  
A:Map position: 13R

Query Match 92.3%; Score 36; DB 2; Length 142;  
Best Local Similarity 87.5%; Pred. No. 13;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKKA 8  
|||||||:  
Db 89 KKKKKKS 96

RESULT 7  
S63193  
Hypothetical protein YNL227c - yeast (*Saccharomyces cerevisiae*)  
N:Alternate names: hypothetical protein N1254  
C:Species: *Saccharomyces cerevisiae*  
C:Date: 27-Apr-1996 #sequence\_revision 03-May-1996 #text\_change 26-Feb-1998  
C:Accession: S63193; S63185; S67371; S72089  
R:Pandolfo, D.; De Antoni, A.; Lanfranchi, G.; Valle, G.  
Submitted to the Protein Sequence Database, April 1996  
A:Reference number: S63188  
A:Accession: S63193  
A:Molecule type: DNA  
A:Residues: 1-590 <PAN>  
A:Cross-references: EMBL:Z71504; NID:g1302248; PID:e239704; PID:g1302250; MIPS:YNL227  
A:Experimental source: strain S288C  
R:Duesterhoeft, A.; Floeth, M.; Fritze, C.; Heuss-Neitzel, D.; Hilbert, H.; Moestl, D.  
Submitted to the Protein Sequence Database, April 1996  
A:Reference number: S62944  
A:Accession: S63185  
A:Molecule type: DNA  
A:Residues: 1-525 <DUE>  
A:Cross-references: EMBL:Z71504; MIPS:YNL227c  
A:Experimental source: strain S288C  
R:Pandolfo, D.; de Antoni, A.; Lanfranchi, G.; Valle, G.  
Submitted to the EMBL Data Library, February 1996  
A:Description: DNA sequence of cosmid 14-5 from chromosome XIV.  
A:Reference number: S67355  
A:Accession: S67371

A:Molecule type: DNA  
A:Residues: 1-590 <PAW>  
A:Cross-references: EMBL:269381; NID:g1183970; PID:e221957; PID:g1183987  
R:Pandolfo, D.; de Antoni, A.; Lanfranchi, G.; Valle, G.  
Yeast 12, 1071-1076, 1996  
A>Title: The DNA sequence of cosmid 14-5 from chromosome XIV reveals 21 open reading frames  
A:Reference number: S72073  
A:Accession: S72089  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-590 <PAF>  
A:Cross-references: EMBL:269381; NID:g1183970; PID:e221957; PID:g1183987  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1996  
C:Genetics:  
A:Map position: 14L  
A>Note: YML227c  
C:Superfamily: dnaJ amino-terminal homology  
F:4-70/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 92.3%; Score 36; DB 2; Length 590;  
Best Local Similarity 87.5%; Pred. No. 42;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKRKA 8  
|:|||||  
Db 460 KKKKKRKA 467

RESULT 8  
S54625  
Probable membrane protein YOR019w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein O2631; hypothetical protein YOL303.8  
C:Species: Saccharomyces cerevisiae  
C>Date: 08-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 12-Dec-1997  
C:Accession: S54625; S66885  
R:de Haan, M.; Maarse, A.C.; Grievell, L.A.  
submitted to the EMBL Data Library, May 1995  
A:Reference number: S54617  
A:Accession: S54625  
A:Molecule type: DNA  
A:Residues: 1-730 <DEH>  
A:Cross-references: EMBL:X87331; NID:g1041652; PID:g829129  
R:de Haan, M.; Grievell, L.A.; Maarse, A.C.  
submitted to the Protein Sequence Database, July 1996  
A:Reference number: S66877  
A:Accession: S66885  
A:Molecule type: DNA  
A:Residues: 1-730 <DEM>  
A:Cross-references: EMBL:Z74927; NID:g1420122; PID:e252323; PID:g1420123; MIPS:YOR019w  
A:Experimental source: strain S288C  
C:Genetics:  
A:Map position: 15R  
C:Keywords: Transmembrane protein  
F:317-333/Domain: transmembrane #status predicted <TMM>

Query Match 92.3%; Score 36; DB 2; Length 730;  
Best Local Similarity 87.5%; Pred. No. 50;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKRKA 8  
|:|||||  
Db 716 KKKKKRKA 723

RESULT 9  
JQ1320  
high mobility group protein Pf16 - Plasmodium falciparum  
C:Species: Plasmodium falciparum  
C>Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 20-Mar-1998  
C:Accession: JQ1320  
R:Gunataka, R.V.; Kandala, J.C.; Reddy, V.D.

Biochem. Biophys. Res. Commun. 182, 412-419, 1992  
A>Title: Cloning and characterization of a highly conserved HMG-like protein (PF16) g  
A:Reference number: JQ1320; MUID:92118046  
A:Accession: JQ1320

A:Molecule type: DNA  
A:Residues: 1-147 <GUN>  
A:Cross-references: GB:M86518; NID:g160325; PID:g160326  
A:Experimental source: strain FCF/3  
C:Comment: This protein interacts with other nuclear proteins and serve as a transcri  
C:Genetics:  
A:Gene: PF16  
F:7-18,19-30/Region: duplication  
F:70-91/Region: aspartic acid/glutamic acid-rich  
F:126-133/Region: basic

Query Match 92.3%; Score 36; DB 2; Length 147;  
Best Local Similarity 87.5%; Pred. No. 13;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKRKA 8  
|:|||||  
Db 38 KKKKKRKA 45

RESULT 10  
I38182  
transcription termination factor TTF-I - human  
C:Species: Homo sapiens (man)  
C>Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 12-Jul-1996  
C:Accession: I38182; S52260  
R:Evans, R.; Grummt, I.  
Proc. Natl. Acad. Sci. U.S.A. 92, 5827-5831, 1995  
A>Title: Molecular coevolution of mammalian ribosomal gene terminator sequences and t  
A:Reference number: I38182; MUID:95320168  
A:Accession: I38182  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-886 <RES>  
A:Cross-references: EMBL:X83973; NID:g639692; PID:g639693

Query Match 92.3%; Score 36; DB 2; Length 886;  
Best Local Similarity 87.5%; Pred. No. 58;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKRKA 8  
|:|||||  
Db 276 KKKKKRKA 283

RESULT 11  
JC5630  
TCOF1 protein - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 14-Oct-1997 #sequence\_revision 07-Nov-1997 #text\_change 21-Nov-1997  
C:Accession: JC5630  
R:Paznekas, W.A.; Zhang, N.; Gridley, T.; Jabs, E.W.  
Biochem. Biophys. Res. Commun. 238, 1-6, 1997  
A>Title: Mouse TCOF1 is expressed widely, has motifs conserved in nucleolar phosphopr  
A:Reference number: JC5630  
A:Accession: JC5630  
A>Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-1320 <PAZ>  
A:Cross-references: DDBJ:AF001794; NID:g2109458; PID:g2109459  
C:Comment: This protein is a nucleolar phosphoprotein with 82 potential phosphorylati  
ing.

Query Match 92.3%; Score 36; DB 2; Length 1320;  
Best Local Similarity 87.5%; Pred. No. 81;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 8  
|||||  
Db 1308 KKKKKKS 1315

## RESULT 12

DPHM2  
N:Altenate names: p67; translation initiation factor eIF-2-associated protein  
C:Species: Homo sapiens (man)  
C>Date: 14-Jul-1995 #sequence\_revision 23-Aug-1996 #text\_change 06-Sep-1996  
C:Accession: S52112  
R:Li, X.; Chang, Y.H.  
Biochim. Biophys. Acta 1260, 333-336, 1995  
A:Title: Molecular cloning of a human complementary DNA encoding an initiation factor 2-  
A:Reference number: S52112  
A:Accession: S52112  
A:Molecule type: mRNA  
A:Residues: 1-478 <LIX>  
C:Cross-references: GB:U13261; NID:g687242; PID:g687243  
C:Genetics:  
A:Gene: GDB:P67EIF2  
A:Cross-references: GDB:S12821  
C:Function:  
A:Description: catalyzes hydrolysis of amino-terminal methionine from proteins  
C:Superfamily: human methionyl aminopeptidase  
C:Keywords: alpha-aminoacylpeptide hydrolase; cobalt; metalloprotein; protein biosynthesis  
F:231,262,459/Binding site: cobalt 2 (Asp, Asp, Glu) #status predicted  
F:262,331,364,459/Binding site: cobalt 1 (Asp, His, Glu, Glu) #status predicted

Query Match 89.7%; Score 35; DB 1; Length 478;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKK 7  
|||||  
Db 98 KKKKKK 104

## RESULT 13

A25910  
small nuclear ribonucleoprotein U2B' - human  
N:Altenate names: snRNP U2B'  
C:Species: Homo sapiens (man)  
C>Date: 16-Aug-1988 #sequence\_revision 02-Dec-1994 #text\_change 20-Mar-1998  
C:Accession: A25910  
R:Heberts, W.J.; Sillescu, P.T.G.; Hoet, M.H.; Schalken, J.A.; Roebroek, A.J.M.; Leuniss  
Proc. Natl. Acad. Sci. U.S.A. 84, 2421-2425, 1987  
A:Title: Analysis of a cDNA clone expressing a human autoimmune antigen: full-length seq  
A:Reference number: A25910; MUID:87175685  
A:Accession: A25910  
A:Molecule type: mRNA  
A:Residues: 1-225 <HAB>  
A:Cross-references: GB:M15841; NID:g340104; PID:g340105  
C:Genetics:  
A:Gene: GDB:SNRNP2  
A:Cross-references: GDB:I25281  
A:Map position: 20pter-20qter  
C:Superfamily: small nuclear ribonucleoprotein U1A; ribonucleoprotein repeat homology  
C:Keywords: duplication; pre-mRNA splicing; RNA binding  
F:8-76/Domain: ribonucleoprotein repeat homology <RRM1>  
F:152-214/Domain: ribonucleoprotein repeat homology <RRM2>

Query Match 89.7%; Score 35; DB 1; Length 225;  
Best Local Similarity 87.5%; Pred. No. 27;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 8  
|||||  
Db 103 KKKKKKA 110

## RESULT 14

PMBYH  
Potassium transport protein TRK1, high affinity - yeast (Saccharomyces cerevisiae)  
N:Altenate names: protein J0693; protein YJL129c  
C:Species: Saccharomyces cerevisiae  
C>Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 12-Dec-1997  
C:Accession: S05849; S56910  
R:Gaber, R.F.; Styles, C.A.; Fink, G.R.  
Mol. Cell. Biol. 8, 2848-2859, 1988  
A:Title: TRK1 encodes a plasma membrane protein required for high-affinity potassium  
A:Reference number: S05849; MUID:88302204  
A:Accession: S05849  
A:Molecule type: DNA  
A:Residues: 1-1235 <GAB>  
A:Cross-references: EMBL:M21328; NID:g171803; PID:g171804  
R:Chiepluch, C.; Korde, E.; Pujol, A.; Jauniaux, J.C.  
submitted to the Protein Sequence Database, September 1995  
A:Reference number: S56891  
A:Accession: S56910  
A:Molecule type: DNA  
A:Residues: 1-1235 <CZ1>  
A:Cross-references: EMBL:249404; NID:g1008329; PID:g1008330; MIPS:YJL129c  
C:Genetics:  
A:Gene: SGD:TRK1  
A:Cross-references: SGD:S000365; MIPS:YJL129c  
A:Map position: 10L  
C:Superfamily: potassium transport protein TRK1  
C:Keywords: ion transport; potassium transport; transmembrane protein  
F:50-70/Domain: transmembrane #status predicted <TM01>  
F:72-96/Domain: transmembrane #status predicted <TM02>  
F:107-127/Domain: transmembrane #status predicted <TM03>  
F:178-800/Domain: transmembrane #status predicted <TM04>  
F:812-834/Domain: transmembrane #status predicted <TM05>  
F:839-859/Domain: transmembrane #status predicted <TM06>  
F:862-882/Domain: transmembrane #status predicted <TM07>  
F:898-918/Domain: transmembrane #status predicted <TM08>  
F:923-943/Domain: transmembrane #status predicted <TM09>  
F:972-991/Domain: transmembrane #status predicted <TM10>  
F:1079-1096/Domain: transmembrane #status predicted <TM11>  
F:1111-1132/Domain: transmembrane #status predicted <TM12>  
F:738/Binding site: ATP (Lys) #status predicted

Query Match 89.7%; Score 35; DB 1; Length 1235;  
Best Local Similarity 100.0%; Pred. No. 11e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKK 7  
|||||  
Db 1052 KKKKKK 1058

## RESULT 15

A46702  
methionyl aminopeptidase (EC 3.4.11.18) 2 - rat  
N:Altenate names: p67; translation initiation factor eIF2-associated protein  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 08-Sep-1997  
C:Accession: A46702  
R:Wu, S.; Gupta, S.; Chatterjee, N.; Hilleman, R.E.; Kinzy, T.G.; Denslow, N.D.; Merril  
J. Biol. Chem. 268, 10796-10801, 1993  
A:Title: Cloning and characterization of complementary DNA encoding the eukaryotic in  
A:Reference number: A46702  
A:Accession: A46702  
A:Molecule type: mRNA  
A:Residues: 1-480 <WDA>  
A:Cross-references: GB:L10652; NID:g204003; PID:g204004  
C:Superfamily: human methionyl aminopeptidase  
C:Keywords: alpha-aminoacylpeptide hydrolase; glycoprotein

Query Match 89.7%; Score 35; DB 2; Length 480;  
Best Local Similarity .100.0%; Pred. No. 50;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KKKKKK 7  
| | | | |  
Db 98 KKKKKK 104

Search completed: July 18, 1999, 06:07:03  
Job time: 249 sec



OM protein - protein search, using sw model

Run on: July 18, 1999, 11:37:49 ; Search time 31.53 Seconds  
(without alignments)  
6.810 Million cell updates/sec

```
Title: US-09-142-043-2
Perfect score: 39
Sequence: 1 KKKKKKA 8
```

Scoring table: BLOSUM62

Searched: 74019 seqs, 26840295 residues

Database : SwissProt\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	39	100.0	437	1	EF1E_XENLA	Q91375 xenopus laae
2	39	100.0	1647	1	SN24_HUMAN	P51532 homo sapien
3	36	92.3	421	1	SN24_YEAST	P38114 saccharomy
4	36	92.3	370	1	CTPT_PLAIF	P49567 plasmodium
5	36	92.3	310	1	IF2B_SCHPO	P56329 schizosacch
6	36	92.3	509	1	NA57_RAT	P40615 rattus norv
7	36	92.3	1085	1	YAFR_SCHPO	P00983 schizosacch
8	36	92.3	142	1	YMH8_YEAST	P03525 saccharomy
9	36	92.3	590	1	YNM7_YEAST	P53683 saccharomy
10	35	89.7	478	1	AMP2_HUMAN	P50579 homo sapien
11	35	89.7	478	1	AMP2_MOUSE	P00863 mus musculu
12	35	89.7	478	1	AMP2_RAT	P38062 rattus norv
13	35	89.7	445	1	ESAL_YEAST	P00869 saccharomy
14	35	89.7	784	1	GCF_HUMAN	P16383 homo sapien
15	35	89.7	312	1	IF2B_DROME	P41375 drosophila
16	35	89.7	333	1	IF2B_HUMAN	P20042 homo sapien
17	35	89.7	333	1	IF2B_RABIT	P41035 cryptotagus
18	35	89.7	270	1	IF2B_WHEAT	Q24473 triticum ae
19	35	89.7	344	1	PAS1_MOUSE	P50580 mus musculu
20	35	89.7	1886	1	POL_COVAY	P50580 mus musculu
21	35	89.7	1257	1	RBBI_HUMAN	P19199 commelina y
22	35	89.7	418	1	RMS1_HUMAN	P29374 homo sapien
23	35	89.7	225	1	RU2B_HUMAN	P49646 homo sapien
24	35	89.7	1395	1	SP41_YEAST	P08579 homo sapien
25	35	89.7	482	1	SR41_YEAST	P38904 saccharomy
26	35	89.7	670	1	SR32_STNY	P74414 synchocystis
27	35	89.7	1411	1	TCOF_HUMAN	P37331 canis famli
28	35	89.7	1393	1	TOP2_PLAIF	Q13428 homo sapien
29	35	89.7	1235	1	TRK1_YEAST	P41001 plasmodium
30	35	89.7	224	1	Y364_MCGE	P12685 saccharomy
31	35	89.7	506	1	YCX2_EDGGR	P47604 mycoplasma
32	35	89.7	547	1	YDRD_SCHPO	P31821 euglena gra
33	35	89.7	153	1	YED3_YEAST	P31821 schizosacch
34	35	89.7	144	1	YHP5_YEAST	P32633 saccharomy
35	35	89.7	167	1	YK20_YEAST	P38808 saccharomy
36	35	89.7	130	1	YNO3_YEAST	P36133 saccharomy
37	35	89.7	91	1	YVDB_YMCC	P53308 saccharomy
38	34	87.2	533	1	YG55_YEAST	P20347 vacuolins v
39	34	87.2	1288	1	YKDI_CAEEL	P53331 saccharomy
40	34	87.2	486	1	YZVL_CAEEL	Q03560 caenorhabdi
41	33	84.6	505	1	ER60_BOVIN	Q21272 caenorhabdi
42	33	84.6	505	1	ER60_HUMAN	P38657 bos taurus
43	33	84.6	504	1	ER60_MOUSE	P30101 homo sapien
						mus musculu

44	33	84.6	505	1	ER60_RAT	PI1598	rattus norv
45	33	84.6	198	1	RAC3_ARATH	G28912	arabidopsis

## ALIGNMENTS

RESULT	1
EF1L_XENLA	
ID	EF1L_XENLA
AC	091375;
DT	15-JUL-1998 (REL. 36, CREATED)
DT	15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT	15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE	ELONGATION FACTOR 1-GAMMA TYPE 2 (EF-1-GAMMA) (P47).
OS	XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC	CHORDATA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AMPHIBIA; ANURA
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=OOCYTE;
RX	MEDLINE: 94155465.
RA	MORALES J., BASEZ T., CORMIER P., MÜLLNER-LORILLON O., BELLE R.,
RA	OSBORNE H.B.;
RL	DEV. GENET. 14:440-448(1993).
CC	-1- FUNCTION: PROBABLY PLAYS A ROLE IN ANCHORING THE COMPLEX TO OTHER
CC	CELLULAR COMPONENTS.
CC	-1- SUBUNIT: EF-1 IS COMPOSED OF FOUR SUBUNITS: ALPHA, BETA,
CC	DELTA, AND GAMMA.
CC	-1- SIMILARITY: TO OTHER SPECIES' GAMMA SUBUNIT AS WELL AS TO THE
CC	N-TERMINAL OF SOME EUKARYOTIC AMINO ACID-tRNA SYNTHETASES.
EMBL	S69726; G54507; -
DR	PROTEIN: P550040; EFIG: 1.
KW	ELONGATION FACTOR; PROTEIN BIOSYNTHESIS.
QO	SEQUENCE 437 AA; 50248 MW; 3D632E6E CRC32.

Query Match	Score	DB	Length
100.0%	39	1	437

Best Local Similarity 100.0%; Pred. NO. 7.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY	1	KKKKKKKA	8
Db	244	KKKKKKKA	251

RESULT 2

ID	SN24_HUMAN	STANDARD;	PRT;	1647	AA
----	------------	-----------	------	------	----

DT	01-OCT-1996 (REL. 34, CREATED)
DT	01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

DI 13-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

DE PROTEIN HOMOLOG 1). (BRAHMA

OS HOMO SAPIENS (HUMAN).

OC EUTHERIA; PRIMATES.

RP SEQUENCE FROM N.A.  
BY MEDLINE: 94050144

KA KHAVARI P.A., PETERSON C.L., TAMKUN J.W., MENDEL D.B., CRABTREE G.R.,  
BL NATURE 366:170-174(1993)

KN [2]  
RP REVISIONS.

RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.

RP SEQUENCE FROM N.A.

RA CHIBA H., MURAMATSU M., NOMOTO A., KATO H.,  
BI NUCLEIC ACIDS RES. 33:1015-1020(2005)

CC -1- FUNCTION: TRANSCRIPTIONAL COACTIVATOR COOPERATING WITH NUCLEAR  
CC HORMONE RECEPTORS TO POTENTIATE TRANSCRIPTIONAL ACTIVATION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.  
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.  
DR EMBL: U29175; G902046; -;  
DR PROSITE: PS00633; BROMODOMAIN\_1; 1.  
DR PROSITE: PS50014; BROMODOMAIN\_2; 1.  
KW TRANSCRIPTION REGULATION; NUCLEAR PROTEIN; ACTIVATOR; BROMODOMAIN;  
FT ATP-BINDING: HELICASE.  
FT DOMAIN 578 588 POLY-LYS.  
FT DOMAIN 663 672 POLY-GLU.  
FT NP\_BIND 779 786 ATP (POTENTIAL).  
FT SITE 881 884 DEGH BOX.  
FT DOMAIN 1360 1364 POLY-GLU.  
FT DOMAIN 1477 1547 BROMODOMAIN.  
FT DOMAIN 1571 1584 POLY-GLU.  
SQ SEQUENCE 1647 AA; 184585 MW; 6D22A3A CRC32;

Query Match 100.0%; Score 39; DB 1; Length 1647;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKKA 8  
|||||||

DB 582 KKKKKKA 589

RESULT 3  
AMP2\_YEAST  
ID AMP2\_YEAST STANDARD; PRT; 421 AA.  
AC P38174;  
DT 01-OCT-1994 (REL. 30, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE METHIONINE AMINOPEPTIDASE 2 (EC 3.4.11.18) (METAP 2) (PEPTIDASE M 2).  
GN MAP2 OR YBL091C OR YBL0701.  
OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMETES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=YPH500;  
RX MEDLINE: 96109265.  
RA LI X., CHANG Y.-H.;  
RN PROC. NATL. ACAD. SCI. U.S.A. 92:12357-12361(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C;  
RX MEDLINE: 96076635.  
RA OBERMAIER B., GASENHUBER J., PIRAVANDI E., DOMDEY H.;  
RN YEAST 11:1103-1112(1995).  
CC -1- FUNCTION: REMOVES THE AMINO-TERMINAL METHIONINE FROM NASCENT  
CC PROTEINS.  
CC -1- CATALYTIC ACTIVITY: L-METHIONYLPEPTIDE + H(2)O -> L-METHIONINE  
CC + PEPTIDE.  
CC -1- COFACTOR: COBALT (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M24C; ALSO KNOWN AS THE  
CC MAP FAMILY 2.  
DR EMBL: U17437; G1045302; -;  
DR EMBL: X79489; G496684; ALT\_INT.  
DR EMBL: Z35852; E304508; -;  
DR PIR: S45411; S45411.  
DR SGD: L0003044; MAP2.  
DR PROSITE: PS01202; MAP\_2; 1.  
KW HYDROLASE; AMINOPEPTIDASE; COBALT.  
FT METAL 194 194 COBALT (BY SIMILARITY).  
FT METAL 205 205 COBALT (BY SIMILARITY).  
FT METAL 274 274 COBALT (BY SIMILARITY).  
FT METAL 307 307 COBALT (BY SIMILARITY).  
FT METAL 403 403 COBALT (BY SIMILARITY).  
FT CONFLICT 86 D -> V (IN REF. 2).

SQ SEQUENCE 421 AA; 47518 MW; 361C0A4E CRC32;

Query Match 92.3%; Score 36; DB 1; Length 421;  
Best Local Similarity 87.5%; Pred. No. 21;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKKA 8  
|||||||

DB 46 KKKKKKS 53

RESULT 4  
CTPT\_PLAFK  
ID CTPT\_PLAFK STANDARD; PRT; 370 AA.  
AC P49587;  
DT 01-FEB-1996 (REL. 33, CREATED)  
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE CHOLINEPHOSPHATE CYTIDYLYLTRANSFERASE (EC 2.7.7.15) (PHOSPHORYLCHOLINE  
DE TRANSFERASE) (CT).  
GN CTPT.  
OS PLASMODIUM FALCIPARUM (ISOLATE K1 / THAILAND).  
OC EUKARYOTA; PROTOZOA; APICOMPLEXA; SPOROZOA; COCCIDIA; EUCCOCCIDIIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 96061933.  
RA YEO H.-J., SRI WIDADA J., MERCEREAU-PUJALON O., VIAL H.J.;  
RN EUR. J. BIOCHEM. 233:62-72(1995).  
CC -1- FUNCTION: CONTROLS PHOSPHATIDYLCHOLINE SYNTHESIS.  
CC -1- CATALYTIC ACTIVITY: CTP + CHOLINE PHOSPHATE -> PYROPHOSPHATE + CDP  
CC CHOLINE.  
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE ASEXUAL INTRACARYOTIC  
CC STAGES.  
CC -1- SIMILARITY: BELONGS TO THE CHOLINEPHOSPHATE CYTIDYLYLTRANSFERASE  
CC FAMILY.  
DR EMBL: X84041; G1054827; -;  
KW TRANSFERASE; NUCLEOTIDYLTRANSFERASE; PHOSPHOLIPID BIOSYNTHESIS.  
FT DOMAIN 57 64 POLY-ASP.  
FT DOMAIN 193 203 POLY-LYS.  
FT DOMAIN 89 268 CATALYTIC (POTENTIAL).  
SQ SEQUENCE 370 AA; 42630 MW; 8C89A8A5 CRC32;

Query Match 92.3%; Score 36; DB 1; Length 370;  
Best Local Similarity 87.5%; Pred. No. 18;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKKA 8  
|||||||

DB 193 KKKKKKS 200

RESULT 5  
IF2B\_SCHPO  
ID IF2B\_SCHPO STANDARD; PRT; 310 AA.  
AC P56329;  
DT 15-JUL-1998 (REL. 36, CREATED)  
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE PROBABLE EUKARYOTIC TRANSLATION INITIATION FACTOR 2 BETA SUBUNIT (EIF-  
DE 2-BETA) (FRAGMENT).  
GN TIE212 OR SPAC6B12.17G.  
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMETES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RA GENTLES S., CHURCHER C.M., BARRELL B.G., RAINDREAN M.A., WOOD V.;  
RN SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- FUNCTION: EIF-2 FUNCTIONS IN THE EARLY STEPS OF PROTEIN SYNTHESIS  
CC BY FORMING A TERNARY COMPLEX WITH GTP AND INITIATOR TRNA. THIS  
CC COMPLEX BINDS TO A 40S RIBOSOMAL SUBUNIT, FOLLOWED BY MRNA BINDING



CC TO FORM A 43S PREINITIATION COMPLEX. JUNCTION OF THE 60S RIBOSOMAL  
CC SUBUNIT TO FORM THE 80S INITIATION COMPLEX IS PRECEDED BY  
CC HYDROLYSIS OF THE GTP BOUND TO EIF-2 AND RELEASE OF AN EIF-2-GDP  
CC BINARY COMPLEX. IN ORDER FOR EIF-2 TO RECYCLE AND CATALYZE ANOTHER  
CC ROUND OF INITIATION, THE GDP BOUND TO EIF-2 MUST EXCHANGE WITH GTP  
CC BY WAY OF A REACTION CATALYZED BY EIF-2B (BY SIMILARITY).  
CC -1- SUBUNIT: HETEROTRIMER COMPOSED OF AN ALPHA, A BETA, AND A GAMMA  
CC CHAIN.  
CC  
CC -1- SIMILARITY: TO OTHER SPECIES EIF-2-BETA AND TO N-TERMINAL OF P67.  
DR EMBL: Z98531; E334107; -  
KM INITIATION FACTOR; PROTEIN BIOSYNTHESIS; ZINC-FINGER.  
FT  
FT NON-TER 1 1  
FT DOMAIN 9 15 POLY-LYS.  
FT DOMAIN 73 80 POLY-LYS.  
FT DOMAIN 105 112 POLY-LYS.  
FT ZN-FING 262 286 POTENTIAL.  
SO SEQUENCE 310 AA; 34743 MW; 3BE1FA9 CRC32;

Query Match 92.3%; Score 36; DB 1; Length 310;  
Best Local Similarity 87.5%; Pred. No. 16;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 8  
DB 106 KKKKKKS 113

RESULT 6  
ID NAS7\_RAT STANDARD; PRT; 509 AA.  
AC P40615;  
DT 01-FEB-1995 (REL. 31, CREATED)  
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE NUCLEOLAR PROTEIN NAS7.  
OS RATTUS NORVEGICUS (RAT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RX MEDLINE: 95096165.  
RA MEIER U., BLOBEL G.;  
RL J. CELL BIOL. 127:1505-1514(1994).  
RN [2]  
RP REVISIONS TO C-TERMINAL.  
RA MEIER U.;  
RL SUBMITTED (DEC-1997) TO THE SWISS-PROT DATA BANK.  
CC -1- SUBUNIT: ASSOCIATES WITH NOP140.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLAR.  
CC -1- SIMILARITY: BELONGS TO THE NAS7/CBF5/TRUB FAMILY.  
DR EMBL: Z24922; G550293; -  
KW NUCLEOLAR PROTEIN.  
FT  
FT CONFLICT 459 509  
SQ SEQUENCE 509 AA; 56615 MW; 643BFF0F CRC32;  
AAPTTPRYKKKKKKKADGGEAAEDGDGDKRRKKK  
KARAAELSG -> RRPDPKRG (IN REF. 1).

Query Match 92.3%; Score 36; DB 1; Length 509;  
Best Local Similarity 87.5%; Pred. No. 24;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 8  
DB 494 KKKKKKA 501

RESULT 7  
ID YAF\_A\_SCHPO STANDARD; PRT; 1085 AA.  
AC 009863;  
DT 01-FEB-1996 (REL. 33, CREATED)

DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 122.9 KD PROTEIN C29E6.10C IN CHROMOSOME I.  
GN SPAC29E6.10C.  
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RA JONES L., MURPHY L., MCNEIL A., SIMPSON I., HARRIS D., BARRELL B.G.,  
RA RAJANDREAM M.A., WALSH S.V.;  
RL SUBMITTED (OCT-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- SIMILARITY: TO YEAST YNL091W.  
DR EMBL: Z66525; G1044936; -  
KM HYPOTHETICAL PROTEIN.  
FT DOMAIN 33 46 POLY-LYS.  
FT DOMAIN 184 194 POLY-LYS.  
FT DOMAIN 695 698 POLY-LYS.  
FT DOMAIN 714 717 POLY-GLU.  
FT DOMAIN 718 721 POLY-GLU.  
FT DOMAIN 938 941 POLY-ARG.  
FT DOMAIN 938 941 POLY-SER.  
SO SEQUENCE 1085 AA; 122940 MW; A95F271 CRC32;

Query Match 92.3%; Score 36; DB 1; Length 1085;  
Best Local Similarity 87.5%; Pred. No. 48;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 8  
DB 186 KKKKKKS 193

RESULT 8  
ID YMBL\_YEAST STANDARD; PRT; 142 AA.  
AC 003525;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 16.2 KD PROTEIN IN PRP24-RRN9 INTERGENIC REGION.  
GN YMR269W OR YMB156.11.  
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RA LYE G., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A.;  
RL SUBMITTED (MAY-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: Z49260; G809092; -  
KW HYPOTHETICAL PROTEIN.  
FT  
FT DOMAIN 61 64 POLY-ARG.  
FT DOMAIN 89 98 POLY-LYS.  
SQ SEQUENCE 142 AA; 16155 MW; 747C7BF8 CRC32;

Query Match 92.3%; Score 36; DB 1; Length 142;  
Best Local Similarity 87.5%; Pred. No. 7.9;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 8  
DB 89 KKKKKKS 96

RESULT 9  
ID YMW7\_YEAST STANDARD; PRT; 590 AA.  
AC P53863;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 68.8 KD PROTEIN IN URE2-SU72 INTERGENIC REGION.

GN YNL227C OR N1254.  
 OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 97051596.  
 RA PANDOLFO D., DE ANTONI A., LANFRANCHI G., VALLE G.;  
 RN YEAST 12:1071-1076(1996).  
 RP [2]  
 RA SEQUENCE OF 1-525 FROM N.A.  
 RA DUSTERHOEFT A., FLOETH M., FRITZ C., HEUSS-NETZEL D., HILBERT H.,  
 RA MOSTEL D.;  
 RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
 CC -1- SIMILARITY: CONTAINS A DNAJ-LIKE DOMAIN.  
 DR EMBL; 269381; E221957; -.  
 DR EMBL; 271504; E239704; -.  
 DR PROSITE; PS00636; DNAJ\_1; 1.  
 DR PROSITE; PS50076; DNAJ\_2; 1.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2; 2.  
 KW HYPOTHETICAL PROTEIN; CHAPERONE; ZINC-FINGER; REPEAT; METAL-BINDING;  
 KW NUCLEAR PROTEIN.  
 FT DOMAIN 3 72 DNAJ-LIKE.  
 FT ZN\_FING 338 362 C2H2-TYPE.  
 FT ZN\_FING 549 573 C2H2-TYPE.  
 SQ SEQUENCE 590 AA; 68782 MW; 26C2A0FD CRC32;

Query Match 92.3%; Score 36; DB 1; Length 590;  
 Best Local Similarity 87.5%; Pred. No. 28;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKKA 8  
 Db 460 KKKKKKA 467

RESULT 10  
 AMP2\_HUMAN  
 ID AMP2\_HUMAN STANDARD; PRT; 478 AA.  
 AC P50579;  
 DT 01-OCT-1996 (REL. 34, CREATED)  
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE METHIONINE AMINOPEPTIDASE 2 (EC 3.4.11.18) (METAP 2) (PEPTIDASE M 2)  
 DE (INITIATION FACTOR 2 ASSOCIATED 67 KD GLYCOPROTEIN) (P67).  
 GN P67EIF2.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP TISSUE-BRAIN.  
 RX MEDLINE; 95372350.  
 RA ARTIN S.M., KENDALL R.L., HALL L., WEAVER L.H., STEWART A.E.,  
 RA MARTIN S.B.W., BRADSHAW R.A.;  
 RN PROC. NATL. ACAD. SCI. U.S.A. 92:7714-7718(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RP TISSUE-LIVER.  
 RX MEDLINE; 95178556.  
 RA LI X., CHANG Y.;  
 RL BIOCHIM. BIOPHYS. ACTA 1260:333-336(1995).  
 CC -1- FUNCTION: REMOVES THE AMINO-TERMINAL METHIONINE FROM NASCENT  
 CC PROTEINS.  
 CC -1- CATALYTIC ACTIVITY: L-METHIONYLPEPTIDE + H(2)O = L-METHIONINE  
 CC + PEPTIDE.  
 CC -1- COFACTOR: COBALT.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M24C; ALSO KNOWN AS THE  
 CC MAP FAMILY 2.  
 DR EMBL; U29607; G903982; -.  
 DR EMBL; U13261; G687243; -.  
 DR MM; 601870; -.

DR PROSITE; PS01202; MAP\_2; 1.  
 KW HYDROLASE; AMINOPEPTIDASE; COBALT.  
 FT DOMAIN 36 46 ARG/LYS-RICH (BASIC).  
 FT DOMAIN 82 93 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 98 106 POLY-LYS.  
 FT METAL 251 251 COBALT (BY SIMILARITY).  
 FT METAL 262 262 COBALT (BY SIMILARITY).  
 FT METAL 331 331 COBALT (BY SIMILARITY).  
 FT METAL 364 364 COBALT (BY SIMILARITY).  
 FT METAL 460 460 COBALT (BY SIMILARITY).  
 SQ SEQUENCE 478 AA; 52891 MW; 18AF9E0A CRC32;

Query Match 89.7%; Score 35; DB 1; Length 478;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKK 7  
 Db 98 KKKKKK 104

RESULT 11  
 AMP2\_MOUSE  
 ID AMP2\_MOUSE STANDARD; PRT; 478 AA.  
 AC O08663;  
 DT 15-JUL-1998 (REL. 36, CREATED)  
 DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
 DE METHIONINE AMINOPEPTIDASE 2 (EC 3.4.11.18) (METAP 2) (PEPTIDASE M 2)  
 DE (INITIATION FACTOR 2 ASSOCIATED 67 KD GLYCOPROTEIN) (P67).  
 GN P67EIF2.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN-C57BL/6 X CBA; TISSUE-LIVER;  
 RA SERIGUCHI S., SUZUKI E.;  
 RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: REMOVES THE AMINO-TERMINAL METHIONINE FROM NASCENT  
 CC PROTEINS (BY SIMILARITY).  
 CC -1- FUNCTION: PROTECTS BY EIF-2 ALPHA-SUBUNIT FROM INHIBITORY  
 CC PHOSPHORYLATION BY EIF-2 KINASES. PLAYS A CRITICAL ROLE IN THE  
 CC REGULATION OF PROTEIN SYNTHESIS. IT ALSO INTERACTS WITH THE  
 CC EIF-2 GAMMA-SUBUNIT (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: L-METHIONYLPEPTIDE + H(2)O = L-METHIONINE  
 CC + PEPTIDE.  
 CC -1- COFACTOR: COBALT (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M24C; ALSO KNOWN AS THE  
 CC MAP FAMILY 2.  
 DR EMBL; AB003144; D1020571; -.  
 DR PROSITE; PS01202; MAP\_2; 1.  
 KW HYDROLASE; AMINOPEPTIDASE; COBALT.  
 FT DOMAIN 36 46 ARG/LYS-RICH (BASIC).  
 FT DOMAIN 82 93 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 98 106 POLY-LYS.  
 FT METAL 251 251 COBALT (BY SIMILARITY).  
 FT METAL 262 262 COBALT (BY SIMILARITY).  
 FT METAL 331 331 COBALT (BY SIMILARITY).  
 FT METAL 364 364 COBALT (BY SIMILARITY).  
 FT METAL 460 460 COBALT (BY SIMILARITY).  
 SQ SEQUENCE 478 AA; 52921 MW; 119C0D5 CRC32;

Query Match 89.7%; Score 35; DB 1; Length 478;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKK 7  
 Db 98 KKKKKK 104

RESULT 12  
AMP2\_RAT STANDARD; PRT; 478 AA.  
ID AMP2\_RAT  
AC P38062;  
DT 01-OCT-1994 (REL. 30, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE METHIONINE AMINOPEPTIDASE 2 (EC 3.4.11.18) (METAP 2) (PEPTIDASE M 2)  
GN (INITIATION FACTOR 2 ASSOCIATED 67 KD GLYCOPROTEIN) (P67).  
OS PE7EUF2.  
GN RATUS NORVEGICUS (RAT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-REUBER H35; TISSUE-LIVER;  
RX MEDLINE: 93266517.  
RA WU S., GUPTA S., CHATTERJEE N., HITEMAN R.E., KINZY T.G.,  
RA DENSILOW N.D., MERRICK W.C., CHAKRABARTI D., OSTERMAN J.C., GUPTA N.K.;  
RL J. BIOL. CHEM. 268:10796-10801(1993).  
RN [2]  
RP REVISIONS TO C-TERMINAL.  
RX MEDLINE: 95372350.  
RA ARFIN S.M., KENDALL R.L., HALL L., WEAVER L.H., STEWART A.E.,  
RA MATTHEWS B.W., BRADSHAW R.A.;  
RL PROC. NATL. ACAD. SCI. U.S.A. 92:7714-7718(1995).  
CC -1- FUNCTION: REMOVES THE AMINO-TERMINAL METHIONINE FROM NASCENT  
CC PROTEINS.  
CC -1- FUNCTION: PROTECTS EIF-2 ALPHA-SUBUNIT FROM INHIBITORY  
CC PHOSPHORYLATION BY EIF-2 KINASES. PLAYS A CRITICAL ROLE IN THE  
CC REGULATION OF PROTEIN SYNTHESIS. IT ALSO INTERACTS WITH THE  
CC EIF-2 GAMMA-SUBUNIT.  
CC -1- CATALYTIC ACTIVITY: L-METHIONYLPEPTIDE + H(2)O = L-METHIONINE  
CC + PEPTIDE.  
CC -1- COFACTOR: COBALF (BY SIMILARITY).  
CC -1- PTM: CONTAINS 12 O-LINKED GLCNAC.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M24C; ALSO KNOWN AS THE  
CC MAP FAMILY 2.  
DR EMBL: L10652; G204004; -  
DR PIR: A46702; A46702.  
DR PROSITE: PS01202; MAP\_2; 1.  
DR HYDROLASE; AMINOPEPTIDASE; COBALF; GLYCOPROTEIN.  
KW DOMAIN 36 46 ARG/LYS-RICH (BASIC).  
FT FT 82 93 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 98 106 POLY-LYS.  
FT METAL 251 251 COBALF (BY SIMILARITY).  
FT METAL 262 262 COBALF (BY SIMILARITY).  
FT METAL 331 331 COBALF (BY SIMILARITY).  
FT METAL 364 364 COBALF (BY SIMILARITY).  
FT METAL 460 460 COBALF (BY SIMILARITY).  
FT METAL 478 478 LRPCKEVSRGDDY -> CAQVYKLSAEKTIKT (IN  
FT CONFLICT 464 478 REF. 1).  
SQ SEQUENCE 478 AA; 53052 MW; 1AA3A160 CRC32;  
Query Match 89.7%; Score 35; DB 1; Length 478;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GN ESAL OR YOR244W OR 05257.  
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C / FY1679;  
RX MEDLINE: 97127829.  
RA BOYER J., MICHAUX G., FAIRHEAD C., GAILLON L., DUJON B.;  
RL YEAST 12:1575-1586(1996).  
CC -1- SIMILARITY: BELONGS TO THE MYST (SAS/MOZ) FAMILY.  
DR EMBL: 275152; E252101; -  
DR SGD: L0003952; ESAL1.  
FT SIMILAR 302 325 TO SOME ACETYLTRANSFERASES.  
FT DOMAIN 91 97 POLY-LYS.  
SQ SEQUENCE 445 AA; 52613 MW; CB2BD89B CRC32;  
Query Match 89.7%; Score 35; DB 1; Length 445;  
Best Local Similarity 87.5%; Pred. No. 31;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
DB 91 KKKKKKA 98  
QY 1 KKKKKKA 8  
DB 91 KKKKKKA 98  
RESULT 14  
GCF\_HUMAN STANDARD; PRT; 784 AA.  
ID GCF\_HUMAN  
AC P16383;  
DT 01-AUG-1990 (REL. 15, CREATED)  
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)  
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)  
DE GC-RICH SEQUENCE DNA-BINDING FACTOR (GCF) (TRANSCRIPTION FACTOR 9)  
DE (TCF-9).  
GN TCF9.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 90075226.  
RA KAGEYAMA R., PASTAN I.;  
RL CELL 59:815-825(1989).  
CC -1- FUNCTION: FACTOR THAT REPRESSSES TRANSCRIPTION. IT BINDS TO THE  
CC GC-RICH SEQUENCES (GCGGGCC) PRESENT IN THE EPIDERMAL GROWTH FACTOR  
CC RECEPTOR, BETA-ACTIN, AND CALCIUM-DEPENDENT PROTEASE PROMOTERS.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED IN HUMAN TISSUES AND CELL  
CC LINES.  
DR EMBL: M29204; G179412; -  
DR PIR: A33633; A33633.  
DR TRANSFAC: T00320; -  
DR MIM: 189901; -  
KW TRANSCRIPTION REGULATION; REPRESSOR; DNA-BINDING; TRANS-ACTING FACTOR;  
KW NUCLEAR PROTEIN.  
FT FT 1 78  
FT DOMAIN 1 86 ARG/LYS-RICH (BASIC).  
FT DOMAIN 23 29 POLY-LYS.  
FT DOMAIN 186 230 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 359 380 LEUCINE-ZIPPER.  
FT DOMAIN 719 740 LEUCINE-ZIPPER.  
SQ SEQUENCE 784 AA; 91001 MW; FB96504B CRC32;  
Query Match 89.7%; Score 35; DB 1; Length 784;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KKKKKKA 7  
DB 23 KKKKKKA 29

```

RESULT 15
IF2B_DROME
ID IF2B_DROME STANDARD: PRT: 312 AA.
AC P41375.
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE EUKARYOTIC TRANSLATION INITIATION FACTOR 2 BETA SUBUNIT (EIF-2-BETA).
GN EIF2B OR EIF2-BETA.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN-OREGON-R;
CC MEDLINE: 94252579.
CC YE X. CAVENER D.R.;
CC GENE 142:271-274(1994).
CC -I- FUNCTION: EIF-2 FUNCTIONS IN THE EARLY STEPS OF PROTEIN SYNTHESIS
CC BY FORMING A TERNARY COMPLEX WITH GTP AND INITIATOR TRNA. THIS
CC PREINITIATION COMPLEX MEDIATES RIBOSOMAL RECOGNITION OF A START
CC CODON DURING THE SCANNING PROCESS OF THE LEADER REGION.
CC -I- SUBUNIT: HETEROTRIMER COMPOSED OF AN ALPHA, A BETA, AND A GAMMA
CC CHAIN.
CC -I- SIMILARITY: TO OTHER SPECIES EIF-2-BETA AND TO N-TERMINAL OF P67.
DR EMBL: L19197; G304676;
DR FLYBASE: FBgn0004926; Elf2-beta.
KW INITIATION FACTOR; PROTEIN BIOSYNTHESIS; ZINC-FINGER.
FT DOMAIN 13 19 POLY-LYS.
FT DOMAIN 63 70 POLY-LYS.
FT DOMAIN 93 99 POLY-GLU.
FT DOMAIN 110 118 POLY-LYS.
FT ZN_FING 260 284 POTENTIAL.
SO SEQUENCE 312 AA; 35217 MW; 57EBAC5 CRC32;

Query Match 89.7%; Score 35; DB 1; Length 312;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKK 7
   111111
Db 13 KKKKKK 19

```

Search completed: July 18, 1999, 11:37:50  
 Job time: 1749 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 06:02:48 ; Search time 63.5 seconds

(without alignments)  
6.950 Million cell updates/sec

Title: US-09-142-043-2

Perfect score: 39  
Sequence: 1 KKKKKKA 8

Scoring table: BLOSUM62

Searched: 180763 segs, 55169189 residues

Database :

SPREMBL\_8:\*  
1: sp\_fungi:\*  
2: sp\_human:\*  
3: sp\_invertebrate:\*  
4: sp\_mammal:\*  
5: sp\_mhc:\*  
6: sp\_organelle:\*  
7: sp\_phage:\*  
8: sp\_plant:\*  
9: sp\_bacteria:\*  
10: sp\_rodent:\*  
11: sp\_virus:\*  
12: sp\_vertibrate:\*  
13: sp\_unclassified:\*  
14: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	39	100.0	2437	3	077393	077393 gallus galli
2	39	100.0	1630	12	090753	090753 mus musculus
3	36	92.3	285	1	012515	012515 saccharomyc
4	36	92.3	426	1	060085	060085 schizosacch
5	36	92.3	730	1	099248	099248 saccharomyc
6	36	92.3	72	1	P89493	P89493 saccharomyc
7	36	92.3	886	2	015361	015361 homo sapien
8	36	92.3	876	2	043840	043840 homo sapien
9	36	92.3	1225	2	000379	000379 homo sapien
10	36	92.3	147	3	025870	025870 plasmodium
11	36	92.3	861	3	021571	021571 caenorhabdi
12	36	92.3	368	3	045198	045198 caenorhabdi
13	36	92.3	557	3	022243	022243 caenorhabdi
14	36	92.3	617	3	044570	044570 caenorhabdi
15	36	92.3	1336	8	081074	081074 arabidopsis
16	36	92.3	1302	10	008857	008857 mus musculu
17	36	92.3	1320	10	008784	008784 mus musculu
18	36	92.3	264	10	035116	035116 rattus norv
19	36	92.3	1247	10	035927	035927 mus musculu
20	35	89.7	1051	1	001694	001694 pneumocysti
21	35	89.7	1435	1	003291	003291 pneumocysti
22	35	89.7	54	1	007058	007058 saccharomyc
23	35	89.7	790	1	002744	002744 pneumocysti
24	35	89.7	246	1	003973	003973 saccharomyc
25	35	89.7	407	2	000536	000536 homo sapien
26	35	89.7	1411	2	099860	099860 homo sapien
27	35	89.7	510	2	015446	015446 homo sapien
28	35	89.7	920	2	015991	015991 homo sapien
29	35	89.7	866	2	015992	015992 homo sapien

## ALIGNMENTS

30	35	89.7	851	2	015993	015993 homo sapien
31	35	89.7	109	2	014039	014039 homo sapien
32	35	89.7	610	2	092732	092732 homo sapien
33	35	89.7	1211	2	099569	099569 homo sapien
34	35	89.7	1035	2	015153	015153 homo sapien
35	35	89.7	394	2	043846	043846 homo sapien
36	35	89.7	1533	2	075560	075560 homo sapien
37	35	89.7	671	2	076094	076094 homo sapien
38	35	89.7	638	3	024259	024259 drosophila
39	35	89.7	771	3	046307	046307 drosophila
40	35	89.7	254	3	061095	061095 trypanosoma
41	35	89.7	261	3	018662	018662 plasmodium
42	35	89.7	46	3	018050	018050 caenorhabdi
43	35	89.7	755	3	020760	020760 caenorhabdi
44	35	89.7	681	3	021572	021572 caenorhabdi
45	35	89.7	357	9	066697	066697 aquifex aeo

RESULT 1  
ID 077393 PRELIMINARY; PRT; 2437 AA.  
AC 077393;  
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE MAL3P6.2 PROTEIN.  
GN MAL3P6.2.  
OS PLASMODIUM FALCIPARUM.  
OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3D7.  
RA MURPHY L., LAWSON D., BARRELL B.;  
RL SUBMITTED (OCT-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; Z98551; E1331931; -;  
SQ SEQUENCE 2437 AA; 295348 MW; E6280382 CRC32;

Query Match 100.0%; Score 39; DB 3; Length 2437;  
Best Local Similarity 100.0%; Pred.No. 3.4e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KKKKKKA 8  
DB 1390 KKKKKKA 1397  
RESULT 2  
ID 090753 PRELIMINARY; PRT; 1630 AA.  
AC 090753;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE BRG1 PROTEIN.  
GN BRG1.  
OS GALLUS GALUS (CHICKEN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;  
OC NEOGNATHAE; GALLIFORMES; PHASINIDAE; PHASININAE; GALLUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=HAEMOPOIETIC;  
RA MEDLINE; 97169142.  
RA GOODWIN G.H.;  
\*Isolation of cDNAs encoding chicken homologues of the yeast SNF2 and  
RT Drosophila Brahma proteins.\*;  
RL GENE 184:27-32(1997).  
DR EMBL; X91637; G996018; -;  
DR PROSITE; P500633; BROMODOMAIN\_1; 1.  
DR PFM; PF00176; SNF2\_N; 1.

DR PFAM; PF00271; helicase\_C; 1.  
DR PFAM; PF00439; bromodomain; 1.  
SQ SEQUENCE 1630 AA; 183420 MW; 2F089105 CRC32;

Query Match 100.0%; Score 39; DB 12; Length 1630;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 8  
Db 580 KKKKKKA 587

## RESULT 3

ID 012515 PRELIMINARY; PRT; 295 AA.  
AC 012515;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
DE ORF YDL173W.  
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCCHAROMYCETALES;  
OC SACCCHAROMYCETACEAE; SACCCHAROMYCES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA POHL T.M.;  
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA MIPS;  
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ALPHAS288C;  
RA POHL T.M.;  
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; Z74221; E253081; -;  
DR EMBL; Z67750; G1061265; -;  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 295 AA; 31886 MW; F9388DD3 CRC32;

Query Match 92.3%; Score 36; DB 1; Length 295;  
Best Local Similarity 87.5%; Pred. No. 1.3e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 8  
Db 277 KKKKKKS 284.

## RESULT 4

ID 060085 PRELIMINARY; PRT; 426 AA.  
AC 060085;  
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE MERIONINE AMINOPEPTIDASE.  
GN SPIC148.03.  
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;  
OC SCHIZOSACCHAROMYCETES; SCHIZOSACCHAROMYCETACEAE;  
OC SCHIZOSACCHAROMYCES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-972H-;  
RA LYNE M., RAINDREAM M.A., BARRELL B.G., VOICAKERT G.;  
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; AL022305; E1285395; -;  
DR PROSITE; PS01202; MAP\_2; 1.

KW AMINOPEPTIDASE.  
SQ SEQUENCE 426 AA; 47271 MW; A80E0013 CRC32;

Query Match 92.3%; Score 36; DB 1; Length 426;  
Best Local Similarity 87.5%; Pred. No. 1.8e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 8  
Db 48 KKKKKKS 55

## RESULT 5

ID 099248 PRELIMINARY; PRT; 730 AA.  
AC 099248;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1996 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE CHROMOSOME XV READING FRAME ORF YOR019W.  
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCCHAROMYCETALES;  
OC SACCCHAROMYCETACEAE; SACCCHAROMYCES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA DE HAAN M., GRIEVEL L.A., MAARSE A.C.;  
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA MIPS;  
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-FY1679;  
RA DE HAAN M., MAARSE A.C., GRIEVEL L.A.;  
RL SUBMITTED (MAY-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN-FY1679;  
RX MEDLINE; 94019318.  
RA DUOMONT M.E., SCHLICHTER J.B., CARDILLO T.S., HAYES M.K., BETHLENDY G.,  
RA SHERMAN F.;  
RT "CYC2 encodes a factor involved in mitochondrial import of yeast  
RT cytochrome C.";  
RT MOL. CELL. BIOL. 13:6442-6451(1993).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN-FY1679;  
RX MEDLINE; 94169519.  
RA LEE Y.S., SHIMIZU J., YODA K., YAMASAKI M.;  
RT "Molecular cloning of a gene, DHS1, which complements a  
RT drug-hypersensitive mutation of the yeast Saccharomyces cerevisiae.";  
RL BIOSCI. BIOTECHNOL. BIOCHEM. 58:391-395(1994).  
DR EMBL; Z74927; E252323; -;  
DR EMBL; X87331; G829129; -;  
SQ SEQUENCE 730 AA; 83365 MW; 09D72CD9 CRC32;

Query Match 92.3%; Score 36; DB 1; Length 730;  
Best Local Similarity 87.5%; Pred. No. 2.9e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 8  
Db 716 KKKKKKS 723

## RESULT 6

ID P89493 PRELIMINARY; PRT; 72 AA.  
AC P89493;  
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)

DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)  
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
ORF YBL091C (FRAGMENT).  
MAP2.  
OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA, FUNGI; ASCOMYCOTA; HEMIASCOCETES; SACCAROMYCETALES;  
SACCAROMYCETACEAE; SACCAROMYCES.  
RN (1)  
RP SEQUENCE FROM N.A.  
RA DOMDEY H., GASSENHUBER H., OBERMAIER B., PIRAVANDI E.;  
RL SUBMITTED (SEP-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN (2)  
RP SEQUENCE FROM N.A.  
RA MIPS;  
RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN (3)  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 95112788.  
RA FELDMAN H., ALJINOVIC G., ANDRE B., BACLET M.C., BARTHE C.,  
BAUR A., BECAM A.M., BITEAU N., BOLES E., BRANDT T., BRENDL M.,  
BRUCKNER M., BUSSEBAU F., CHRISTANGEN C., CONTRERAS R., CROUZET M.,  
CZIEPLUCH C., DEMOLIS N., DELAVEAU T., DOIGNON F., DOMDEY H.,  
DUESTERER S., DUBOIS E., DUTON B., EL BAKKOURI M., ENTIAN K.D.,  
FEUERMAN M., FIEBS W., FOBO G.M., FRITZ C., GASSENHUBER H.,  
GLASDORFF N., GOFEAU A., GRIVELL L.A., DE HAAN M., HEIN C.,  
HERBERT C.J., HOLLENBERG C.P., HOLMSTROM K., JACO C., JACQUET M.,  
JAUNIAUX J.C., JONNIAUX J.L., KALLESE T., KIESAU P., KIRCHRATH L.,  
KOETTER P., KOROL S., LIEBL S., LOGCHE M., LOHAN A.J.E., LOUIS E.J.,  
LI Z.Y., MAT M.J., MALLETT L., MANNHART G., MESSENGUY F., MIOSGA T.,  
MOLEMAN F., MOELLER S., NASR F., OBERMAIER B., PEREA J., PIERARD A.,  
PIRAVANDI E., POHL F.M., POHL T.M., POTIER S., PROTT M., PURELLE B.,  
RAMZANI RAD M., RIEGER M., ROSE M., SCHAFF-GERTSCHLAGER I.,  
SCHRENS B., SCHWARZLOSE C., SKALA J., SLOMINSKI P.P., SMITS P.H.M.,  
SOUCIET J.L., STEENSA H.Y., STUCKA R., URESTIAZU A.,  
VAN DER AART Q.J., VAN DYCK L., VASSAKOTTI A., VETTER I.,  
VIENENDELS F., VISSERS S., WAGNER G., DE WENGIOSE P., WOLFE K.H.,  
ZAGULSKI K., ZIMMERMANN F.K., MEWES H., KLEINE K.;  
RA "Complete DNA sequence of yeast chromosome II.";  
RT EMO J. 13:5795-5809(1994).  
DR EMBL; E304571; -.  
FT NON\_TER 72 72  
SQ SEQUENCE 72 AA; 8174 MW; E3699D12 CRC32;

Query Match 92.3%; Score 36; DB 1; Length 72;  
Best Local Similarity 87.5%; Pred. No. 38;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 8  
ID 1111111;  
DB 46 KKKKKKS 53

RESULT 7  
015361 PRELIMINARY; PRT; 886 AA.  
AC 015361;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE TRANSCRIPTION FACTOR.  
GN TTF-1.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA, METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; HOMO.  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 95320168.  
RA EVERS R., GRUMET I.;  
RT "Molecular coevolution of mammalian ribosomal gene terminator  
sequences and the transcription termination factor TTF-1.";  
RL PROC. NATL. ACAD. SCI. U.S.A. 92:5827-5831(1995).  
DR EMBL; X83973; G639693; -.

KW TRANSCRIPTION TERMINATION.  
SQ SEQUENCE 886 AA; 101218 MW; 93D6AEF5 CRC32;

Query Match 92.3%; Score 36; DB 2; Length 886;  
Best Local Similarity 87.5%; Pred. No. 3.5e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 8  
ID 1111111;  
DB 276 KKKKKKS 283

RESULT 8  
043840 PRELIMINARY; PRT; 876 AA.  
AC 043840;  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE NEURAL PLAKOPHILIN RELATED ARM-REPEAT PROTEIN (FRAGMENT).  
GN NRPAP.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA, METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; HOMO.  
RN (1)  
RP SEQUENCE FROM N.A.  
RC TISSUE-BRAIN;  
RX MEDLINE; 98002299.  
RA PAPENHOLZ R., FRANK W.W.;  
RT "Identification and localization of a neurally expressed member of the  
RT plakoglobin/armadillo multigene family.";  
RL DIFFERENTIATION 61:293-304(1997).  
DR EMBL; U52351; G2822195; -.  
FT NON\_TER 1 1  
SQ SEQUENCE 876 AA; 96442 MW; 02D752C6 CRC32;

Query Match 92.3%; Score 36; DB 2; Length 876;  
Best Local Similarity 87.5%; Pred. No. 3.4e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 8  
ID 1111111;  
DB 462 KKKKKKS 469

RESULT 9  
000379 PRELIMINARY; PRT; 1225 AA.  
AC 000379; 013589;  
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE DELTA-CATENIN.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA, METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; HOMO.  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 97366296.  
RA ZHOU J., LIYANAGE U., MEDINA M., HO C., SIMONS A.D., KOSIK K.S.;  
RT "Presentilin 1 interaction in the brain with a novel member of the  
RT Armadillo family.";  
RL NEUROREPORT 8:2085-2090(1997).  
RN (2)  
RP SEQUENCE FROM N.A.  
RC TISSUE-BRAIN;  
RA LU Q., PAREDES M., MEDINA M., ZHOU J., CAVALLO R., PEIFER M.,  
RA ORECCHIO L., KOSIK K.S.;  
RT "d-catenin, an adhesive junction associated protein which promotes  
RT motile behavior.";  
RL SUBMITTED (OCT-1998) TO EMBL/GENBANK/DBJ DATA BANKS.

[3]  
RN SEQUENCE OF 865-1225 FROM N.A.  
RP MEDLINE: 97202103.  
RA STIMONS A.D., OVERHAUSER J., LOVETT M.:  
RT "Isolation of cDNAs from the Crl-du-chat critical region by direct  
screening of a chromosome 5-specific cDNA library."  
RL GENOME RES. 7:118-127(1997).  
DR EMBL: U96136; G3712673; -  
DR EMBL: U52828; G2253589; -  
PR EMBL: PF00514; Armadillo\_seg. 4.  
SQ SEQUENCE 1225 AA; 132664 MW; 903ED1DF CRC32;  
  
Query Match 92.3%; Score 36; DB 2; Length 1225;  
Best Local Similarity 87.5%; Pred. No. 4,6e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 KKKKKKA 8  
Db 811 KKKKKKS 818  
  
RESULT 10  
ID Q25870 PRELIMINARY; PRT; 147 AA.  
AC Q25870;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE HIGH MOBILITY GROUP-LIKE PROTEIN (PS16).  
GN PS16.  
OS PLASMODIUM FALCIPARUM.  
OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 92118046.  
RA GONTANA R.V., KANDALA J.C., REDDY V.D.:  
RT "Cloning and characterization of a highly conserved HMG-like protein  
RT (PF16) gene from Plasmodium falciparum."  
RL BIOCHEM. BIOPHYS. RES. COMMUN. 182:412-419(1992).  
DR EMBL: M6518; G160326; -  
SQ SEQUENCE 147 AA; 17175 MW; AFA6385A CRC32;  
  
Query Match 92.3%; Score 36; DB 3; Length 147;  
Best Local Similarity 87.5%; Pred. No. 72;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 KKKKKKA 8  
Db 38 KKKKKKS 45  
  
RESULT 11  
ID Q21571 PRELIMINARY; PRT; 861 AA.  
AC Q21571;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE M28.9.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;  
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX GARDNER A.:  
RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,  
RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,

RA DEAR S., DU Z., DUREIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,  
RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAM J.,  
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LOYD C.,  
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,  
RA RIEKEN L., ROOPER A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,  
RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,  
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
RA WILKINSON-SPROAT J., WOLDMAN P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans."  
RL NATURE 368:32-38(1994).  
DR EMBL: Z49911; G872082; -  
SQ SEQUENCE 861 AA; 93141 MW; FEE94A0E CRC32;  
  
Query Match 92.3%; Score 36; DB 3; Length 861;  
Best Local Similarity 87.5%; Pred. No. 3,4e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 KKKKKKA 8  
Db 733 KKKKKKS 740  
  
RESULT 12  
ID 045198 PRELIMINARY; PRT; 368 AA.  
AC 045198;  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE W09G12.7 PROTEIN.  
GN W09G12.7.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;  
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-BRISTOL N2;  
RX MEDLINE: 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,  
RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,  
RA DEAR S., DU Z., DUREIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,  
RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAM J.,  
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LOYD C.,  
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,  
RA RIEKEN L., ROOPER A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,  
RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,  
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
RA WILKINSON-SPROAT J., WOLDMAN P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans."  
RL NATURE 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX STRAIN-BRISTOL N2;  
RX BECK C., FRONICK W., WILSON R.;  
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RX STRAIN-BRISTOL N2;  
RX WATERSTON R.;  
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: AF047653; G2911882; -  
SQ SEQUENCE 368 AA; 41823 MW; 5B503425 CRC32;  
  
Query Match 92.3%; Score 36; DB 3; Length 368;  
Best Local Similarity 87.5%; Pred. No. 1,6e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 KKKKKKA 8  
Db 1 KKKKKKA 8



Db 126 KKKKKKS 133

RESULT 13

ID 022243 PRELIMINARY; PRT: 557 AA.

AC 022243; (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DT 01-NOV-1996 (TREMBLREL. 08, LAST SEQUENCE UPDATE)

DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

GN TOC10.3

OS CAENORHABDITIS ELEGANS.

OC EUDAROTIA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;

OC RHABDITINA; RHABDITIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 94150718.

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,

RA BURTON J., CONNELL M., COSEY T., COOPER J., COULSON A., CRAXTON M.,

RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,

RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,

RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,

RA MURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,

RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,

RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,

RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,

RA WILKINSON-SPROAT J., WOHLDMAN P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans";

RL NATURE 368:32-38(1994).

RN [2]

RP SEQUENCE FROM N.A.

RX WU X.;

RA SUBMITTED (DEC-1995) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [3]

RP SEQUENCE FROM N.A.

RL WATERSTON R.;

RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL: U41013; G1086632; -

DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.

DR PROSITE: PS00109; PROTEIN\_KINASE\_TIR; 1.

DR PFAM: PF00017; SH2; 1.

DR PFAM: PF00069; PKINASE; 1.

SQ SEQUENCE 557 AA; 62426 MW; 04B7CF6D CRC32;

Query Match 92.3%; Score 36; DB 3; Length 557;

Best Local Similarity 87.5%; Pred. No. 2.3e+02;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKKA 8

Db 106 KKKKKKS 113

RESULT 14

ID 044570 PRELIMINARY; PRT: 617 AA.

AC 044570;

DT 01-JUN-1998 (TREMBLREL. 06, CREATED)

DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE W03F8.2 PROTEIN.

GN W03F8.2.

OS CAENORHABDITIS ELEGANS.

OC EUDAROTIA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;

OC RHABDITINA; RHABDITIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN-BRISTOL N2;

RX MEDLINE: 94150718.

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,

RA BURTON J., CONNELL M., COSEY T., COOPER J., COULSON A., CRAXTON M.,

RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,

RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,

RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,

RA MURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,

RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,

RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,

RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,

RA WILKINSON-SPROAT J., WOHLDMAN P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans";

RL NATURE 368:32-38(1994).

RN [2]

RP SEQUENCE FROM N.A.

RX STRAIN-BRISTOL N2;

RA JOHNSON D., BRADSHAW H., KEPPLER D.;

RL SUBMITTED (JAN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [3]

RP SEQUENCE FROM N.A.

RX STRAIN-BRISTOL N2;

RA WATERSTON R.;

RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL: AF039041; G2736384; -

SQ SEQUENCE 617 AA; 70526 MW; 6BD02F9A CRC32;

Query Match 92.3%; Score 36; DB 3; Length 617;

Best Local Similarity 87.5%; Pred. No. 2.5e+02;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKKA 8

Db 157 KKKKKKS 164

RESULT 15

ID 081074 PRELIMINARY; PRT: 1336 AA.

AC 081074;

DT 01-NOV-1998 (TREMBLREL. 08, CREATED)

DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE PUTATIVE ZINC-FINGER PROTEIN.

GN T914.14.

OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).

OC EUDAROTIA; VIRIDIPHYTES; CHAROPHYTA/EMBRYOPHYTA GROUP; EMBRYOPHYTA;

OC TRACHEOPHYTA; EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA;

OC EUDICOTYLEDONS; ROSIDAE; CAPRALES; BRASSICACEAE; ARABIDOPSIS.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-CV. COLUMBIA;

RA ROUNSLEY S.D., LIN X., KAUL S., SHEA T.P., FUJII C.Y., NASON T.M.,

RA SHEN M., RONNING C.M., FRASER C.M., SOMERVILLE C.B., VENTER J.C.;

RT SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL: AC005315; G3461846; -

GN ZINC-FINGER.

SQ SEQUENCE 1336 AA; 152069 MW; C781515E CRC32;

Query Match 92.3%; Score 36; DB 8; Length 1336;

Best Local Similarity 87.5%; Pred. No. 5e+02;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKKA 8

Db 943 KKKKKKS 950

Search completed: July 18, 1999, 06:02:49

Job time: 3104 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 00:42:53 ; Search time 58.51 Seconds  
(without alignments)  
25.925 Million cell updates/sec

Title: US-09-142-043-3

Perfect score: 422

Sequence: 1 RIPCGLPPTITNGDFISTNR.....NDQVGIWGPAPQCIIPNK 75

Scoring table: BLOSUM62

Searched: 162890 segs, 20225328 residues

Database: A\_Geneseq\_34:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	422	100.0	2317	P92219	CRI protein New nu
2	422	100.0	2039	R11810	Human complement t
3	422	100.0	543	R28543	CRI-4 (35E, 37Y) a
4	422	100.0	543	R28544	CRI-4 (35E) analog
5	422	100.0	543	R28545	CRI-4 (37I) analog
6	422	100.0	543	R28546	CRI-4 (44T, 47D, 4
7	422	100.0	543	R28547	CRI-4 (52S, 53S, 5
8	422	100.0	543	R28548	CRI-4 (57V, 59K) a
9	422	100.0	543	R28549	CRI-4 (64K, 65T) a
10	422	100.0	543	R28550	CRI-4 (64K) analog
11	422	100.0	543	R28551	CRI-4 (65T) analog
12	422	100.0	543	R28552	CRI-4 (78T, 79D) a
13	422	100.0	543	R28553	CRI-4 (85R, 87N) a
14	422	100.0	543	R28554	CRI-4 (92T, 94H) a
15	422	100.0	543	R28555	CRI-4 (92T) analog
16	422	100.0	543	R28556	CRI-4 (94H) analog
17	422	100.0	543	R28557	CRI-4 (99H, 103E)
18	422	100.0	543	R28558	CRI-4 (109N, 110A,
19	422	100.0	543	R28559	CRI-4 (114-117STR
20	422	100.0	543	R28560	CRI-4 (114S) analo
21	422	100.0	543	R28561	CRI-4 (115T) analo
22	422	100.0	543	R28562	CRI-4 (116K) analo
23	422	100.0	543	R28563	CRI-4 (117P) analo
24	422	100.0	543	R28564	CRI-4 (116K, 117P)
25	422	100.0	543	R28565	CRI-4 (121Q) analo
26	422	100.0	483	R29091	CRI-4 (amino acids
27	422	100.0	481	R29092	CRI-4 (amino acids
28	422	100.0	543	R28566	CRI-4 (318R, 319K)
29	422	100.0	543	R28567	CRI-4 (318-321 RNP
30	422	100.0	543	R28568	CRI-4 (347T, 349Y)
31	422	100.0	543	R28569	CRI-4 (369-376 STR
32	422	100.0	543	R28570	CRI-4 (266-274 KLK
33	422	100.0	543	R28571	CRI-4 (364-367 NMA
34	422	100.0	2039	R35743	CRI. Nucleic acid
35	422	100.0	197	R47152	Sequence of solubl
36	422	100.0	76	R47153	Sequence of solubl
37	422	100.0	254	R47154	Sequence of solubl
38	422	100.0	254	R47155	Sequence of solubl
39	422	100.0	133	R47156	Sequence of solubl
40	422	100.0	211	W45909	SCR 1-3 of complem
41	422	100.0	198	W45897	SCR 1-3 of complem
42	422	100.0	214	W45898	SCR 1-3 of complem
43	422	100.0	214	W45905	SCR 1-3 of complem

## ALIGNMENTS

44	422	100.0	1930	1	W45899	Human complement r
45	422	100.0	778	1	W73147	Amino acid sequenc
ALIGNMENTS						
RESULT 1						
ID	P92219	standard; protein; 2317 AA.				
AC	P92219;					
DE	22-FEB-1990	(first entry)				
DE	CRI protein					
KW	Complement; cofactor.					
OS	Homo sapiens (human).					
FH	key	Location/Qualifiers				
FT	peptide	10..50				
FT		/label= signal_peptide				
PN	W08909220-A.					
PD	05-OCT-1989.					
PF	31-MAR-1989;	U01358.				
PR	01-APR-1988;	US-176532.				
PA	(TCEU) T Cell Sciences Inc; (UYJO) The Johns Hopkins University;					
PA	(BRIG*) The Brigham and Women's Hospital.					
PI	Fearon DT, Klickstein LB, Wong W, Carson G, Concino MF, Makrides SC;					
DR	WPI: 89-309498/42.					
DR	N-PSDB: N91477.					
PT	New nucleic acid sequences encoding new CRI protein - and its fragment,					
PT	for diagnosis and control of complement-related immune defects,					
PT	inflammation, myocardial infarct, etc					
PS	Claim 1; fig. 1; 191pp; English.					
CC	This is full-length CRI protein, and shortened forms are new, lacking					
CC	the transmembrane region. The proteins and fragments bind C3b and/or					
CC	C4b, have cofactor activity and inhibit C3 and C5 convertase activity.					
CC	In the sequence, x-untranslated region. This has 7 short consensus					
CC	repeats (CR) of 4 long direct homologous repeats (LHR) (see fig. 10 in					
CC	patent). Deletion mutants, eg LHR-A for binding of C4b, and LHR-C					
CC	for C3b binding, be made, and LHR-B and -C can detect CR2 sequences.					
CC	They are useful in diagnosing and treating immune disorders, and prevent					
CC	perfusion injury.					
SQ	Sequence 2317 AA;					
Query Match						
		100.0%;	Score 422;	DB 1;	Length 2317;	
		Best Local Similarity	100.0%;	Pred. No. 3.5e-42;		
		Matches 75;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
OY	1	RIPCGLPPTITNGDFISTNR	ENHFGSVVTVRCNPGSGRGKVFELVGEPSICTSNDQV	60		
DB	172	RIPCGLPPTITNGDFISTNR	ENHFGSVVTVRCNPGSGRGKVFELVGEPSICTSNDQV	231		
OY	61	GIWGPAPQCIIPNK	75			
DB	232	GIWGPAPQCIIPNK	246			
RESULT 2						
ID	R11810	standard; protein; 2039 AA.				
AC	R11810;					
DE	25-JUN-1991	(first entry)				
DE	Human complement type 1 receptor.					
KW	Complement system; C3b/C4b receptor; CRI; allergic reaction;					
KW	immune response; clone lambda T109.1.					
OS	Homo sapiens.					
FH	key	Location/Qualifiers				
FT	peptide	1..41				
FT		/label= putative signal peptide				
FT	protein	42..2039				
FT		/label= CRI				
PN	W09105047-A.					
PD	18-APR-1991.					
PF	25-SEP-1990;	U05454.				

26-SEP-1989; US-412745.  
26-SEP-1990; US-912349.  
PA (TCELL-) T CELL SCI INC.  
(UYJO.) JOHNS HOPKINS UNIVERSITY.  
PA (BRIG-) BRIGHAM AND WOMEN'S HOSPITAL.  
PI Fearon DT, Klichslein LB, Wong WM, Carson GR, Hoh M, Concino MF,  
PI Makrides SC, Marsh HC;  
DR WPI: 91-132854/18.  
DR N-PSDB: Q11642.  
PT Human complement receptor type 1 gene, encoded proteins and  
PT fragments - for treatment of immune disorders, myocardial infarct,  
PS damage due to inflammation and in treatment of thrombosis  
PS Claim 41; Fig 1: 234pp; English.  
CC The invention also covers fragments of this protein which have the  
CC ability to bind C3b and/or C4b, have cofactor I activity or can  
CC inhibit C3 or C5 convertase activity. The full-length protein, or  
CC its specified fragments are used to treat patients with immune  
CC disorders or a disorder caused by inappropriate complement  
CC activity. The protein is also used to treat thrombotic conditions  
CC in humans and animals. See also Q11643.  
SO Sequence 2039 AA;

Query Match 100.0%; Score 422; DB 1; Length 2039;  
Best Local Similarity 100.0%; Pred. No. 3e-42;  
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RIPCGLPTTNGDFTSTNRENFHYGSVYTRCNPSSGGRKVFELVGEPSYCTSNDDOV 60  
DB 163 RIPCGLPTTNGDFTSTNRENFHYGSVYTRCNPSSGGRKVFELVGEPSYCTSNDDOV 222  
QY 61 GIWSPAPQCIIPNK 75  
DB 223 GIWSPAPQCIIPNK 237

RESULT 3  
R28543  
ID R28543 standard; peptide; 543 AA.  
AC R28543:  
DT 19-MAR-1993 (first entry)  
DE CRI-4 (35E, 37Y) analogue.  
KW short consensus repeat; regulator of complement activation;  
KW C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT region 1..60  
FT /label= SCR-1  
FT region 61..122  
FT /label= SCR-2  
FT region 451..510  
FT /label= SCR-8  
FT region 511..543  
FT /label= SCR-9  
FT /note= "TRUNCATED"  
FT misc-difference 35  
FT /note= "Gly substituted by Glu (SCR-8)"  
FT misc-difference 37  
FT /note= "Ser substituted by Tyr (SCR-8)"  
PN EP-512733-A.  
PD 11-NOV-1992.  
PF 28-APR-1992; 303826.  
PR 03-MAY-1991; US-695514.  
PA (UNIW ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Kirsch M;  
DR WPI: 92-375009/46.  
PT Complement activity regulator protein analogues - useful for  
PT treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11; Fig 2 and R11810: 23pp; English.  
CC The cDNA clone designated CRI-4 was described in J. Exp. Med. (1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCRs of CRI. The invention concerns analogues of "regulator of

CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed. Positions 35 and  
CC 37 of SCR-1 and the corresponding positions in SCR-8 have been  
CC identified as important in C4b binding. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQ  
CC accession number R11810 and descriptions in the disclosure.  
SO Sequence 543 AA;

Query Match 100.0%; Score 422; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 6.2e-43;  
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RIPCGLPTTNGDFTSTNRENFHYGSVYTRCNPSSGGRKVFELVGEPSYCTSNDDOV 60  
DB 122 RIPCGLPTTNGDFTSTNRENFHYGSVYTRCNPSSGGRKVFELVGEPSYCTSNDDOV 181  
QY 61 GIWSPAPQCIIPNK 75  
DB 182 GIWSPAPQCIIPNK 196

RESULT 4  
R28544  
ID R28544 standard; peptide; 543 AA.  
AC R28544:  
DT 19-MAR-1993 (first entry)  
DE CRI-4 (35E) analogue.  
KW short consensus repeat; regulator of complement activation;  
KW C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT region 1..60  
FT /label= SCR-1  
FT region 61..122  
FT /label= SCR-2  
FT region 451..510  
FT /label= SCR-8  
FT region 511..543  
FT /label= SCR-9  
FT /note= "TRUNCATED"  
FT misc-difference 35  
FT /note= "Gly substituted by Glu (SCR-8)"  
PN EP-512733-A.  
PD 11-NOV-1992.  
PF 28-APR-1992; 303826.  
PR 03-MAY-1991; US-695514.  
PA (UNIW ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Kirsch M;  
DR WPI: 92-375009/46.  
PT Complement activity regulator protein analogues - useful for  
PT treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11; Fig 2 and R11810: 23pp; English.  
CC The cDNA clone designated CRI-4 was described in J. Exp. Med. (1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCRs of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed. Positions 35 and  
CC 37 of SCR-1 and the corresponding positions in SCR-8 have been  
CC identified as important in C4b binding. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQ  
CC accession number R11810 and descriptions in the disclosure.  
SO Sequence 543 AA;

Query Match 100.0%; Score 422; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 6.2e-43;  
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 RIPCGLPPTITNGDFISTNRENHFHGSVVTYRCNPGSGGKRVFELVGEPSICTSNDQV 60
    |||||||
DB 122 RIPCGLPPTITNGDFISTNRENHFHGSVVTYRCNPGSGGKRVFELVGEPSICTSNDQV 181
OY 61 GIMSGPAPQCIIPNK 75
    |||||||
DB 182 GIMSGPAPQCIIPNK 196

RESULT 5
R28545
ID R28545 standard; peptide: 543 AA.
AC R28545;
DT 19-MAR-1993 (first entry)
DE CRI-4 (37Y) analogue.
KW short consensus repeat; regulator of complement activation;
   C3b binding; C4b binding; human complement type 1 receptor.
OS Homo sapiens.
FH Key.
FT Location/Qualifiers
FT 1..60
FT /label= SCR-1
FT region 61..122
FT /label= SCR-2
FT 451..510
FT /label= SCR-8
FT 511..543
FT /label= SCR-9
FT /note= "TRUNCATED"
FT misc_difference 37
FT /note= "Ser substituted by Tyr (SCR-8)"
FT
FT
FT EP-512733-A.
FT PD 11-NOV-1992.
FT PF 28-APR-1992; 303826.
FT PR 03-MAY-1991; US-695514.
FT PA (UNITW ) UNIV WASHINGTON.
FT PI Atkinson JP, Hourcade D, Kiyoch M;
FT DR WPI; 92-375009/46.
FT PT Complement activity regulator protein analogues - useful for
FT treating auto-immune diseases, to suppress transplant rejection,
FT for diagnosis etc.
FT PS Claim 11: Fig 2 and R11810; 23pp; English.
FT CC The CDNA clone designated CRI-4 was described in J.Exp.Med.(1988)
FT CC 168:1255-1270. It encodes the first 8 and a half amino terminal
FT CC SCRs of CRI. The invention concerns analogues of "regulator of
FT CC complement activation" proteins or truncated, hybrid or recombinant
FT CC forms of them. CRI-4 is a preferred truncated form and a number of
FT CC specified substitution variants of it are claimed. Positions 35 and
FT CC 37 of SCR-1 and the corresponding positions in SCR-8 have been
FT CC identified as important in C4b binding. The specification does not
FT CC contain the CRI-4 sequence; the sequence given here was constructed
FT CC from the full-length CRI amino acid sequence having GENESEQ
FT CC accession number R11810 and descriptions in the disclosure.
FT SQ Sequence 543 AA;

Query Match 100.0%; Score 422; DB 1; Length 543;
Best Local Similarity 100.0%; Pred. No. 6.2e-43;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RIPCGLPPTITNGDFISTNRENHFHGSVVTYRCNPGSGGKRVFELVGEPSICTSNDQV 60
    |||||||
DB 122 RIPCGLPPTITNGDFISTNRENHFHGSVVTYRCNPGSGGKRVFELVGEPSICTSNDQV 181
OY 61 GIMSGPAPQCIIPNK 75
    |||||||
DB 182 GIMSGPAPQCIIPNK 196

RESULT 6
R28546
ID R28546 standard; peptide: 543 AA.
AC R28546;
```

```
DT 19-MAR-1993 (first entry)
DE CRI-4 (44T, 47D, 49L) analogue.
KW short consensus repeat; regulator of complement activation;
   C3b binding; C4b binding; human complement type 1 receptor.
OS Homo sapiens.
FH Key.
FT Location/Qualifiers
FT 1..60
FT /label= SCR-1
FT region 61..122
FT /label= SCR-2
FT 451..510
FT /label= SCR-8
FT 511..543
FT /label= SCR-9
FT /note= "TRUNCATED"
FT misc_difference 44
FT /note= "Ile substituted by Thr (SCR-8)"
FT misc_difference 47
FT /note= "Lys substituted by Asp (SCR-8)"
FT misc_difference 49
FT /note= "Ser substituted by Leu (SCR-8)"
FT
FT
FT EP-512733-A.
FT PD 11-NOV-1992.
FT PF 28-APR-1992; 303826.
FT PR 03-MAY-1991; US-695514.
FT PA (UNITW ) UNIV WASHINGTON.
FT PI Atkinson JP, Hourcade D, Kiyoch M;
FT DR WPI; 92-375009/46.
FT PT Complement activity regulator protein analogues - useful for
FT treating auto-immune diseases, to suppress transplant rejection,
FT for diagnosis etc.
FT PS Claim 11: Fig 2 and R11810; 23pp; English.
FT CC The CDNA clone designated CRI-4 was described in J.Exp.Med.(1988)
FT CC 168:1255-1270. It encodes the first 8 and a half amino terminal
FT CC SCRs of CRI. The invention concerns analogues of "regulator of
FT CC complement activation" proteins or truncated, hybrid or recombinant
FT CC forms of them. CRI-4 is a preferred truncated form and a number of
FT CC specified substitution variants of it are claimed in which certain
FT CC positions in SCR-1 which have been identified as important for the
FT CC degree of C3b- and C4b-binding are substituted by amino acids from
FT CC the corresponding positions in SCR-8. The specification does not
FT CC contain the CRI-4 sequence; the sequence given here was constructed
FT CC from the full-length CRI amino acid sequence having GENESEQ
FT CC accession number R11810 and descriptions in the disclosure.
FT SQ Sequence 543 AA;

Query Match 100.0%; Score 422; DB 1; Length 543;
Best Local Similarity 100.0%; Pred. No. 6.2e-43;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RIPCGLPPTITNGDFISTNRENHFHGSVVTYRCNPGSGGKRVFELVGEPSICTSNDQV 60
    |||||||
DB 122 RIPCGLPPTITNGDFISTNRENHFHGSVVTYRCNPGSGGKRVFELVGEPSICTSNDQV 181
OY 61 GIMSGPAPQCIIPNK 75
    |||||||
DB 182 GIMSGPAPQCIIPNK 196

RESULT 7
R28547
ID R28547 standard; peptide: 543 AA.
AC R28547;
DT 19-MAR-1993 (first entry)
DE CRI-4 (52S, 53S, 54P) analogue.
KW short consensus repeat; regulator of complement activation;
   C3b binding; C4b binding; human complement type 1 receptor.
OS Homo sapiens.
FH Key.
FT Location/Qualifiers
FT 1..60
FT /label= SCR-1
FT region 61..122
```

```
FT FT /label=SCR-2
FT region 451..510
FT /label=SCR-8
FT region 511..543
FT /label=SCR-9
FT /note="TRUNCATED"
FT misc_difference 52
FT /note="Thr substituted by Ser (SCR-8)"
FT misc_difference 53
FT /note="Gly substituted by Ser (SCR-8)"
FT misc_difference 54
FT /note="Ala substituted by Pro (SCR-8)"
PN EP-512733-A.
PD 11-NOV-1992.
PE 28-APR-1992: 303826.
PR 03-MAY-1991: US-695514.
PA (UNIW ) UNIV WASHINGTON.
PI Atkinson JP, Hourcade D, Krych M;
DR WPI: 92-375009/46.
PT Complement activity regulator protein analogues - useful for
PT treating auto-immune diseases, to suppress transplant rejection,
PT for diagnosis etc.
PS Claim 11: Fig 2 and R1810: 23pp: English.
CC The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)
CC 168:1255-1270. It encodes the first 8 and a half amino terminal
CC SCRs of CRI. The invention concerns analogues of "regulator of
CC complement activation" proteins or truncated, hybrid or recombinant
CC forms of them. CRI-4 is a preferred truncated form and a number of
CC specified substitution variants of it are claimed in which certain
CC positions in SCR-1 which have been identified as important for the
CC degree of C3b- and C4b-binding are substituted by amino acids from
CC the corresponding positions in SCR-8. The specification does not
CC contain the CRI-4 sequence; the sequence given here was constructed
CC from the full-length CRI amino acid sequence having GENSEQ
CC accession number R1810 and descriptions in the disclosure.
SQ Sequence 543 AA:

Query Match 100.0%; Score 422; DB 1; Length 543;
Best Local Similarity 100.0%; Pred. No. 6.2e-43;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RIPGCLPTTNGDFISTNRENFHSGSVYTYRCNPGSGRKFELVGEPSIYCTSNDDQV 60
DB 122 RIPGCLPTTNGDFISTNRENFHSGSVYTYRCNPGSGRKFELVGEPSIYCTSNDDQV 181
QY 61 GWSGPAPOCIIPNK 75
DB 182 GWSGPAPOCIIPNK 196

RESULT 8
R28548
ID R28548 standard: peptide; 543 AA.
AC R28548;
DT 19-MAR-1993 (first entry)
DE CRI-4 (57V, 59K) analogue.
KW short consensus repeat; regulator of complement activation;
KW C3b binding; C4b binding; human complement type 1 receptor.
OS Homo sapiens.
FH Key
FT Location/Qualifiers
FT 1..60
FT /label=SCR-1
FT /label=SCR-2
FT /label=SCR-8
FT /label=SCR-9
FT /note="TRUNCATED"
FT misc_difference 57
FT /note="Arg substituted by Val (SCR-8)"
FT misc_difference 59
```

```
FT FT /note="Arg substituted by Lys (SCR-8)"
PN EP-512733-A.
PD 11-NOV-1992.
PE 28-APR-1992: 303826.
PR 03-MAY-1991: US-695514.
PA (UNIW ) UNIV WASHINGTON.
PI Atkinson JP, Hourcade D, Krych M;
DR WPI: 92-375009/46.
PT Complement activity regulator protein analogues - useful for
PT treating auto-immune diseases, to suppress transplant rejection,
PT for diagnosis etc.
PS Claim 11: Fig 2 and R1810: 23pp: English.
CC The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)
CC 168:1255-1270. It encodes the first 8 and a half amino terminal
CC SCRs of CRI. The invention concerns analogues of "regulator of
CC complement activation" proteins or truncated, hybrid or recombinant
CC forms of them. CRI-4 is a preferred truncated form and a number of
CC specified substitution variants of it are claimed in which certain
CC positions in SCR-1 which have been identified as important for the
CC degree of C3b- and C4b-binding are substituted by amino acids from
CC the corresponding positions in SCR-8. The specification does not
CC contain the CRI-4 sequence; the sequence given here was constructed
CC from the full-length CRI amino acid sequence having GENSEQ
CC accession number R1810 and descriptions in the disclosure.
SQ Sequence 543 AA:

Query Match 100.0%; Score 422; DB 1; Length 543;
Best Local Similarity 100.0%; Pred. No. 6.2e-43;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RIPGCLPTTNGDFISTNRENFHSGSVYTYRCNPGSGRKFELVGEPSIYCTSNDDQV 60
DB 122 RIPGCLPTTNGDFISTNRENFHSGSVYTYRCNPGSGRKFELVGEPSIYCTSNDDQV 181
QY 61 GWSGPAPOCIIPNK 75
DB 182 GWSGPAPOCIIPNK 196

RESULT 9
R28549
ID R28549 standard: peptide; 543 AA.
AC R28549;
DT 19-MAR-1993 (first entry)
DE CRI-4 (64K, 65T) analogue.
KW short consensus repeat; regulator of complement activation;
KW C3b binding; C4b binding; human complement type 1 receptor.
OS Homo sapiens.
FH Key
FT Location/Qualifiers
FT 1..60
FT /label=SCR-1
FT /label=SCR-2
FT /label=SCR-8
FT /label=SCR-9
FT /note="TRUNCATED"
FT misc_difference 64
FT /note="Arg substituted by Lys (SCR-9)"
FT misc_difference 65
FT /note="Asn substituted by Thr (SCR-9)"
PN EP-512733-A.
PD 11-NOV-1992.
PE 28-APR-1992: 303826.
PR 03-MAY-1991: US-695514.
PA (UNIW ) UNIV WASHINGTON.
PI Atkinson JP, Hourcade D, Krych M;
DR WPI: 92-375009/46.
PT Complement activity regulator protein analogues - useful for
PT treating auto-immune diseases, to suppress transplant rejection,
PT for diagnosis etc.
```

PS Claim 11: Fig 2 and R11810: 23pp; English.  
CC The cDNA clone designated CRI-4 was described in J. Exp. Med. (1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCRs of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-2 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-9. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQ  
CC accession number R11810 and descriptions in the disclosure.  
SQ Sequence 543 AA.

Query Match 100.0%; Score 422; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 6,2e-43;  
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RIPCGLPPTINGDFISTNRENHFHGSVYTYRCNPGSGRKVFELVGEPSITCTSNDDQV 60  
DB 122 RIPCGLPPTINGDFISTNRENHFHGSVYTYRCNPGSGRKVFELVGEPSITCTSNDDQV 181  
OY 61 GWSGAPAPQCIIPNK 75  
DB 182 GWSGAPAPQCIIPNK 196

RESULT 10  
R28550  
ID R28550 standard; peptide: 543 AA.  
AC R28550: (first entry)  
DT 19-MAR-1993  
DE CRI-4 (64k) analogue.  
KW short consensus repeat; regulator of complement activation;  
KW C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT region 1..60  
FT /label= SCR-1  
FT region 61..122  
FT /label= SCR-2  
FT region 451..510  
FT /label= SCR-8  
FT region 511..543  
FT /label= SCR-9  
FT /note= "TRUNCATED"  
FT misc\_difference 64  
FT /note= "Arg substituted by Lys (SCR-9)"

PN EP-512733-A.  
PD 11-NOV-1992.  
PF 28-APR-1992; 303826.  
PR 03-MAY-1991; US-695514.  
PA (UNIW ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Kirsch M;  
DR WPI: 92-375009/46.  
PT Complement activity regulator protein analogues - useful for  
PT treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11: Fig 2 and R11810: 23pp; English.  
CC The cDNA clone designated CRI-4 was described in J. Exp. Med. (1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCRs of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-2 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-9. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQ  
CC accession number R11810 and descriptions in the disclosure.

SQ Sequence 543 AA;

Query Match 100.0%; Score 422; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 6,2e-43;  
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RIPCGLPPTINGDFISTNRENHFHGSVYTYRCNPGSGRKVFELVGEPSITCTSNDDQV 60  
DB 122 RIPCGLPPTINGDFISTNRENHFHGSVYTYRCNPGSGRKVFELVGEPSITCTSNDDQV 181  
OY 61 GWSGAPAPQCIIPNK 75  
DB 182 GWSGAPAPQCIIPNK 196

RESULT 11  
R28551  
ID R28551 standard; peptide: 543 AA.  
AC R28551: (first entry)  
DT 19-MAR-1993  
DE CRI-4 (65T) analogue.  
KW short consensus repeat; regulator of complement activation;  
KW C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT region 1..60  
FT /label= SCR-1  
FT region 61..122  
FT /label= SCR-2  
FT region 451..510  
FT /label= SCR-8  
FT region 511..543  
FT /label= SCR-9  
FT /note= "TRUNCATED"  
FT misc\_difference 64  
FT /note= "Arg substituted by Lys (SCR-9)"  
FT misc\_difference 65  
FT /note= "Asn substituted by Thr (SCR-9)"

PN EP-512733-A.  
PD 11-NOV-1992.  
PF 28-APR-1992; 303826.  
PR 03-MAY-1991; US-695514.  
PA (UNIW ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Kirsch M;  
DR WPI: 92-375009/46.  
PT Complement activity regulator protein analogues - useful for  
PT treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11: Fig 2 and R11810: 23pp; English.  
CC The cDNA clone designated CRI-4 was described in J. Exp. Med. (1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCRs of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-2 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-9. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQ  
CC accession number R11810 and descriptions in the disclosure.

Query Match 100.0%; Score 422; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 6,2e-43;  
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RIPCGLPPTINGDFISTNRENHFHGSVYTYRCNPGSGRKVFELVGEPSITCTSNDDQV 60  
DB 122 RIPCGLPPTINGDFISTNRENHFHGSVYTYRCNPGSGRKVFELVGEPSITCTSNDDQV 181





```

FT      /label= SCR-9
FT      /note= "TRUNCATED"
FT      misc_difference 92
FT      /note= "Lys substituted by Thr (SCR-9)"
FT      /note= "Tyr substituted by His (SCR-9)"
FT      misc_difference 94
FT      /note= "Tyr substituted by His (SCR-9)"
PN      EP-512733-A.
PD      11-NOV-1992.
PF      28-APR-1992; 303826.
PR      03-MAY-1991; US-695514.
PA      (UNIW ) UNIV WASHINGTON.
PI      Atkinson JP, Hourcade D, Krych M.
DR      WPI; 92-375009/46.
PT      Complement activity regulator protein analogues - useful for
PT      treating auto-immune diseases, to suppress transplant rejection,
PT      for diagnosis etc.
PS      Claim 11; Fig 2 and R1810; 23pp; English.
CC      The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)
CC      168:1255-1270. It encodes the first 8 and a half amino terminal
CC      SCRs of CRI. The invention concerns analogues of "regulator of
CC      complement activation" proteins or truncated, hybrid or recombinant
CC      forms of them. CRI-4 is a preferred truncated form and a number of
CC      specified substitution variants of it are claimed in which certain
CC      positions in SCR-2 which have been identified as important for the
CC      degree of C3b- and C4b-binding are substituted by amino acids from
CC      the corresponding positions in SCR-9. The specification does not
CC      contain the CRI-4 sequence; the sequence given here was constructed
CC      from the full-length CRI amino acid sequence having GENESEQ
CC      accession number R1810 and descriptions in the disclosure.
SQ      Sequence 543 AA;

```

```

Query Match      100.0%; Score 422; DB 1; Length 543;
Best Local Similarity 100.0%; Pred. No. 6.2e-43;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 RIPCGLPPTTNGDFISTNRENFHGSVVTYRCNPGSGGKRVFELVGEPSIYCTSNDDQV 60
        |||||||
DB      122 RIPCGLPPTTNGDFISTNRENFHGSVVTYRCNPGSGGKRVFELVGEPSIYCTSNDDQV 181
        |||||||
QY      61 GIWSGAPPOCIIPNK 75
        |||||||
DB      182 GIWSGAPPOCIIPNK 196
        |||||||

```

```

RESULT 15
R28535
ID      R28555 standard; peptide; 543 AA.
AC      R28555;
DT      19-MAR-1993 (first entry)
DE      CRI-4 (927) analogue.
KW      Short consensus repeat; regulator of complement activation;
KW      C3b binding; C4b binding; human complement type 1 receptor.
OS      Homo sapiens.
FH      Key
FT      region      Location/Qualifiers
FT      region      1..60
FT      region      /label= SCR-1
FT      region      61..122
FT      region      /label= SCR-2
FT      region      451..510
FT      region      /label= SCR-8
FT      region      511..543
FT      region      /label= SCR-9
FT      misc_difference 92
FT      /note= "TRUNCATED"
FT      /note= "Lys substituted by Thr (SCR-9)"
PN      EP-512733-A.
PD      11-NOV-1992.
PF      28-APR-1992; 303826.
PR      03-MAY-1991; US-695514.
PA      (UNIW ) UNIV WASHINGTON.
PI      Atkinson JP, Hourcade D, Krych M.
DR      WPI; 92-375009/46.

```

```

PT      Complement activity regulator protein analogues - useful for
PT      treating auto-immune diseases, to suppress transplant rejection,
PT      for diagnosis etc.
PS      Claim 11; Fig 2 and R1810; 23pp; English.
CC      The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)
CC      168:1255-1270. It encodes the first 8 and a half amino terminal
CC      SCRs of CRI. The invention concerns analogues of "regulator of
CC      complement activation" proteins or truncated, hybrid or recombinant
CC      forms of them. CRI-4 is a preferred truncated form and a number of
CC      specified substitution variants of it are claimed in which certain
CC      positions in SCR-2 which have been identified as important for the
CC      degree of C3b- and C4b-binding are substituted by amino acids from
CC      the corresponding positions in SCR-9. The specification does not
CC      contain the CRI-4 sequence; the sequence given here was constructed
CC      from the full-length CRI amino acid sequence having GENESEQ
CC      accession number R1810 and descriptions in the disclosure.
SQ      Sequence 543 AA;

```

```

Query Match      100.0%; Score 422; DB 1; Length 543;
Best Local Similarity 100.0%; Pred. No. 6.2e-43;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

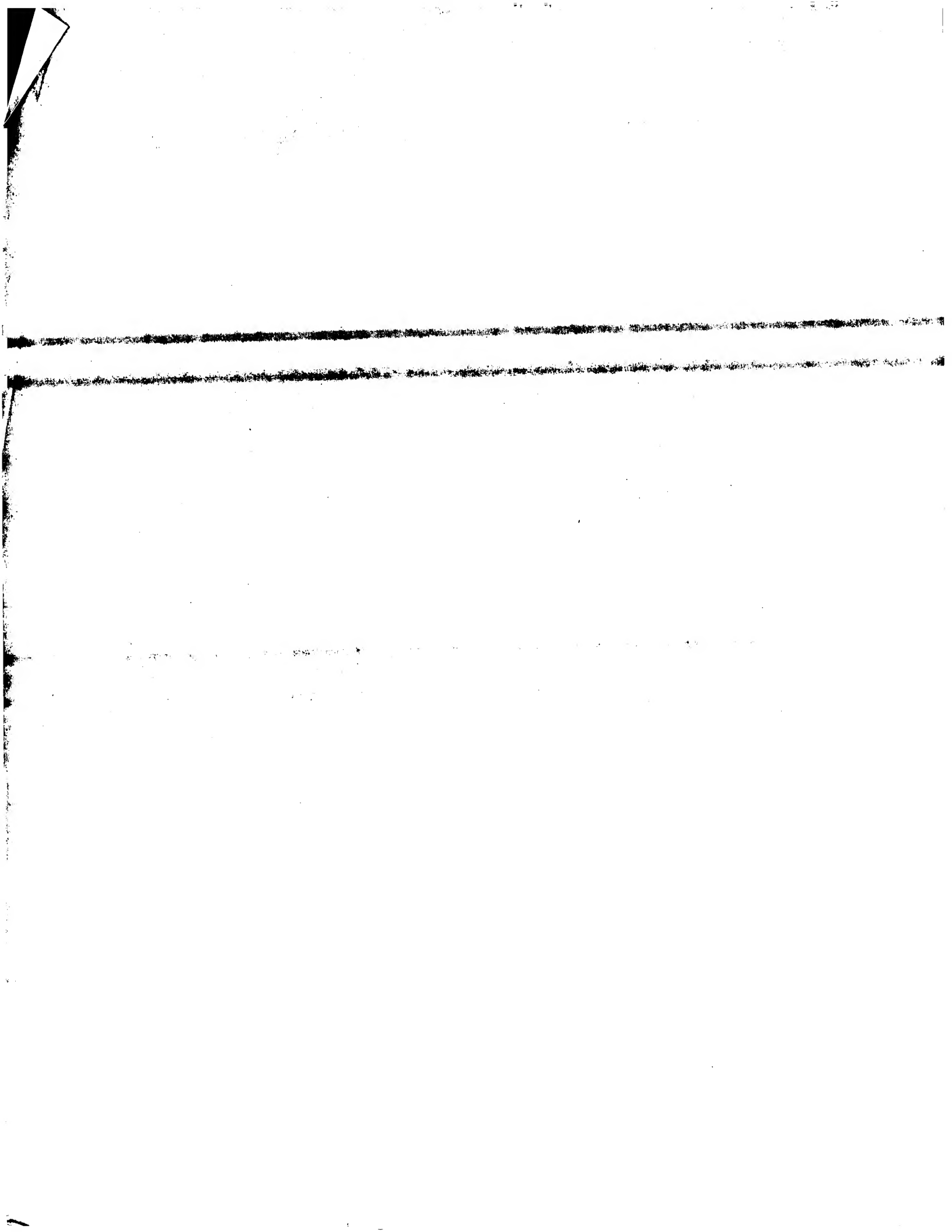
QY      1 RIPCGLPPTTNGDFISTNRENFHGSVVTYRCNPGSGGKRVFELVGEPSIYCTSNDDQV 60
        |||||||
DB      122 RIPCGLPPTTNGDFISTNRENFHGSVVTYRCNPGSGGKRVFELVGEPSIYCTSNDDQV 181
        |||||||
QY      61 GIWSGAPPOCIIPNK 75
        |||||||
DB      182 GIWSGAPPOCIIPNK 196
        |||||||

```

```

Search completed: July 18, 1999, 00:42:53
Job time: 8586 sec

```



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 06:07:55 ; Search time 39.54 Seconds

(without alignments)  
18.718 Million cell updates/sec

Title: US-09-142-043-3

Perfect score: 422  
Sequence: 1 RIPCGLPPTTNGDFISTNR.....NDQVGVSGPAPCIIIPNK 75

Scoring table: BLOSUM62

Searched: 106577 seqs, 9868381 residues

Database : Issued\_Patents\_AA:\*

1: /cgn2\_6/prodata/1/laa/5A.COMB.pep:\*  
2: /cgn2\_6/prodata/1/laa/5B.COMB.pep:\*  
3: /cgn2\_6/prodata/1/laa/PCTUS9.COMB.pep:\*  
4: /cgn2\_6/prodata/1/laa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	422	100.0	197	2	US-08-356-361-27	Sequence 27, Appl
2	422	100.0	76	2	US-08-356-361-28	Sequence 28, Appl
3	422	100.0	254	2	US-08-356-361-29	Sequence 29, Appl
4	422	100.0	254	2	US-08-356-361-30	Sequence 30, Appl
5	422	100.0	133	2	US-08-356-361-31	Sequence 31, Appl
6	422	100.0	197	2	US-08-769-967A-27	Sequence 27, Appl
7	422	100.0	76	2	US-08-769-967A-28	Sequence 28, Appl
8	422	100.0	254	2	US-08-769-967A-29	Sequence 29, Appl
9	422	100.0	254	2	US-08-769-967A-30	Sequence 30, Appl
10	422	100.0	133	2	US-08-769-967A-31	Sequence 31, Appl
11	159	37.7	324	1	US-08-310-416A-14	Sequence 14, Appl
12	159	37.7	324	1	US-08-888-171-14	Sequence 14, Appl
13	159	37.7	323	2	US-08-435-149-2	Sequence 2, Appl
14	159	37.7	577	2	US-08-435-149-3	Sequence 3, Appl
15	132.5	31.4	254	1	US-08-310-416A-13	Sequence 13, Appl
16	132.5	31.4	293	1	US-08-310-416A-16	Sequence 16, Appl
17	132.5	31.4	169	1	US-08-310-416A-18	Sequence 18, Appl
18	132.5	31.4	377	2	US-08-528-057-2	Sequence 2, Appl
19	132.5	31.4	370	2	US-08-528-057-42	Sequence 42, Appl
20	132.5	31.4	373	2	US-08-528-057-44	Sequence 44, Appl
21	132.5	31.4	324	2	US-08-528-057-46	Sequence 46, Appl
22	132.5	31.4	254	2	US-08-888-171-13	Sequence 13, Appl
23	132.5	31.4	293	2	US-08-888-171-16	Sequence 16, Appl
24	132.5	31.4	169	2	US-08-888-171-18	Sequence 18, Appl
25	132.5	31.4	254	2	US-08-435-149-1	Sequence 1, Appl
26	104.5	24.8	263	1	US-07-906-983-2	Sequence 2, Appl
27	86	20.4	1257	1	US-08-340-428B-49	Sequence 49, Appl
28	85	20.1	1083	1	US-08-296-014A-2	Sequence 2, Appl
29	85	20.1	1019	1	US-08-296-014A-4	Sequence 4, Appl
30	85	20.1	1083	2	US-08-596-405-2	Sequence 2, Appl
31	85	20.1	1019	2	US-08-596-405-4	Sequence 4, Appl
32	84	19.3	310	4	5256642-10	Patent No. 5256642
33	84	19.3	310	4	5472939-10	Patent No. 5472939
34	81	19.2	385	2	US-08-340-539A-2	Sequence 2, Appl
35	81	19.2	385	2	US-08-461-592B-2	Sequence 2, Appl
36	81	19.2	372	2	US-08-513-278-2	Sequence 2, Appl
37	81	19.2	933	3	PCT-US93-03837-1	Sequence 1, Appl
38	78	18.5	830	1	US-08-110-158-4	Sequence 4, Appl
39	78	18.5	830	3	PCT-US91-05059-2	Sequence 2, Appl

40	78	18.5	912	3	PCT-US95-03747-2	Sequence 2, Appl
41	77	18.2	340	4	5256642-2	Patent No. 5256642
42	77	18.2	4	4	5472939-2	Patent No. 5472939
43	76.5	18.1	610	1	US-08-365-470-3	Sequence 3, Appl
44	74.5	17.7	62	1	US-08-210-266A-12	Sequence 12, Appl
45	74.5	17.7	62	1	US-08-688-675-12	Sequence 12, Appl

## ALIGNMENTS

```
RESULT 1
US-08-356-361-27
; Sequence 27, Application US/08356361
; Patent No. 5833989
; GENERAL INFORMATION:
; APPLICANT: Smith, Richard A.G.
; APPLICANT: Dodd, Ian
; APPLICANT: Freeman Mary A.
; APPLICANT: Mossakowska, Danuta E.I.
; TITLE OF INVENTION: No. 5833989el Compounds
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham-Corporate Intellectual Property
; STREET: P.O. Box 1539
; CITY: King of Prussia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/356, 361
; APPLICATION NUMBER: US/08/356, 361
; FILING DATE: 03-Jul-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jarvis, Herbert H.
; REGISTRATION NUMBER: 31,171
; REFERENCE/DOCKET NUMBER: P30423
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (610) 270-5019
; TELEFAX: (610) 270-5090
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 197 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: N-terminal
; US-08-356-361-27

Query Match 100.0% Score 422; DB 2; Length 197;
Best Local Similarity 100.0%; Pred. No. 3.7e-43;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RIPCGLPPTTNGDFISTNRNFHGSVVTYRCNPGSGGRKVFELVGEPSICTSNDQV 60
|||
Db 123 RIPCGLPPTTNGDFISTNRNFHGSVVTYRCNPGSGGRKVFELVGEPSICTSNDQV 182
QY 61 GWSGAPAPCIIIPNK 75
|||||
Db 183 GWSGAPAPCIIIPNK 197

RESULT 2
US-08-356-361-28
; Sequence 28, Application US/08356361
; Patent No. 5833989
; GENERAL INFORMATION:
```

APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman Mary A.  
APPLICANT: Mossakowska, Danuta E.I.  
TITLE OF INVENTION: No. 5833989el Compounds  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham-Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,361  
FILING DATE: 03-Jul-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jervis, Herbert H.  
REGISTRATION NUMBER: 31,171  
REFERENCE/DOCKET NUMBER: P30423  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5019  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 76 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-356-361-28

Query Match 100.0%; Score 422; DB 2; Length 76;  
Best Local Similarity 100.0%; Pred. No. 1.2e-43;  
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 RIPCGLPPTTNGDFISTNRENFHYSVYTRCNPBGSGGRKVELGEPSTICTSNDQV 60  
Db 2 RIPCGLPPTTNGDFISTNRENFHYSVYTRCNPBGSGGRKVELGEPSTICTSNDQV 61  
OY 61 GWSGPAQCIIIPNK 75  
Db 62 GWSGPAQCIIIPNK 76  
RESULT 3  
US-08-356-361-29  
Sequence 29, Application US/08356361  
Patent No. 5833989  
GENERAL INFORMATION:  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman Mary A.  
APPLICANT: Mossakowska, Danuta E.I.  
TITLE OF INVENTION: No. 5833989el Compounds  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham-Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,361  
FILING DATE: 03-Jul-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jervis, Herbert H.  
REGISTRATION NUMBER: 31,171  
REFERENCE/DOCKET NUMBER: P30423  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5019  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-356-361-29

Query Match 100.0%; Score 422; DB 2; Length 254;  
Best Local Similarity 100.0%; Pred. No. 5.1e-43;  
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 RIPCGLPPTTNGDFISTNRENFHYSVYTRCNPBGSGGRKVELGEPSTICTSNDQV 60  
Db 123 RIPCGLPPTTNGDFISTNRENFHYSVYTRCNPBGSGGRKVELGEPSTICTSNDQV 182  
OY 61 GWSGPAQCIIIPNK 75  
Db 183 GWSGPAQCIIIPNK 197

RESULT 4  
US-08-356-361-30  
Sequence 30, Application US/08356361  
Patent No. 5833989  
GENERAL INFORMATION:  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman Mary A.  
APPLICANT: Mossakowska, Danuta E.I.  
TITLE OF INVENTION: No. 5833989el Compounds  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham-Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,361  
FILING DATE: 03-Jul-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jervis, Herbert H.  
REGISTRATION NUMBER: 31,171  
REFERENCE/DOCKET NUMBER: P30423  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5019  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-356-361-30

Query Match 100.0%; Score 422; DB 2; Length 254;  
Best Local Similarity 100.0%; Pred. No. 5.1e-43;  
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RIPCGLPPTTNGDFISTNRENFHYGSVVYTRCNPSSGGRKVEFLVGEPSIYCTSDNDQV 60  
|||||  
DB 123 RIPCGLPPTTNGDFISTNRENFHYGSVVYTRCNPSSGGRKVEFLVGEPSIYCTSDNDQV 182

OY 61 GIMSGPAPOCIIPNK 75  
|||||  
DB 183 GIMSGPAPOCIIPNK 197

RESULT 5  
US-08-356-361-31

; Sequence 31, Application US/08356361  
; Patent No. 583989  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Richard A.G.  
; APPLICANT: Dodd, Ian  
; APPLICANT: Freeman Mary A.  
; APPLICANT: Mossakowska, Danuta E.I.  
; TITLE OF INVENTION: No. 583989el Compounds  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham-Corporate Intellectual Property  
; STREET: P.O. Box 1539  
; CITY: King of Prussia  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19406  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/356,361  
; FILING DATE: 03-Jul-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jervis, Herbert H.  
; REGISTRATION NUMBER: 31,171  
; REFERENCE/DOCKET NUMBER: P30423  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (610) 270-5019  
; TELEFAX: (610) 270-5090  
; INFORMATION FOR SEQ ID NO: 31:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 133 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: internal  
US-08-356-361-31

Query Match 100.0%; Score 422; DB 2; Length 133;  
Best Local Similarity 100.0%; Pred. No. 2.3e-43;  
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RIPCGLPPTTNGDFISTNRENFHYGSVVYTRCNPSSGGRKVEFLVGEPSIYCTSDNDQV 60  
|||||  
DB 2 RIPCGLPPTTNGDFISTNRENFHYGSVVYTRCNPSSGGRKVEFLVGEPSIYCTSDNDQV 61  
OY 61 GIMSGPAPOCIIPNK 75

DB 62 GIMSGPAPOCIIPNK 76  
|||||

RESULT 6  
US-08-769-967A-27

; Sequence 27, Application US/08769967A  
; Patent No. 5859223  
; GENERAL INFORMATION:  
; APPLICANT: Mossakowska, Danuta E.I.  
; APPLICANT: Smith, Richard A.G.  
; APPLICANT: Freeman, Anne Mary  
; APPLICANT: Dodd, Ian  
; TITLE OF INVENTION: Soluble CRI Derivatives  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporate Intellectual Property  
; STREET: P.O. Box 1539  
; CITY: King of Prussia  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19406  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/769,967A  
; FILING DATE:  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/440,569  
; FILING DATE: 15-May-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: King, William T.  
; REGISTRATION NUMBER: 30,954  
; REFERENCE/DOCKET NUMBER: P30423C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (610) 270-5364  
; TELEFAX: (610) 270-5090  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 197 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: N-terminal  
US-08-769-967A-27

Query Match 100.0%; Score 422; DB 2; Length 197;  
Best Local Similarity 100.0%; Pred. No. 3.7e-43;  
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RIPCGLPPTTNGDFISTNRENFHYGSVVYTRCNPSSGGRKVEFLVGEPSIYCTSDNDQV 60  
|||||  
DB 123 RIPCGLPPTTNGDFISTNRENFHYGSVVYTRCNPSSGGRKVEFLVGEPSIYCTSDNDQV 182  
OY 61 GIMSGPAPOCIIPNK 75  
|||||  
DB 183 GIMSGPAPOCIIPNK 197

RESULT 7

US-08-769-967A-28  
; Sequence 28, Application US/08769967A  
; Patent No. 5859223  
; GENERAL INFORMATION:  
; APPLICANT: Mossakowska, Danuta E.I.  
; APPLICANT: Smith, Richard A.G.  
; APPLICANT: Dodd, Ian  
; APPLICANT: Freeman, Anne Mary

TITLE OF INVENTION: Soluble CRI Derivatives  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,967A  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,569  
FILING DATE: 15-May-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: P30423C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5364  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 76 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-769-967A-28

Query Match 100.0%; Score 422; DB 2; Length 76;  
Best Local Similarity 100.0%; Pred. No. 1.2e-43;  
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RIPGCLPPTINGDFISTNRENFHYGSVVYTRCNPBGSGGRKVFELVGPSTICTSNDQV 60  
DB 2 RIPGCLPPTINGDFISTNRENFHYGSVVYTRCNPBGSGGRKVFELVGPSTICTSNDQV 61  
QY 61 GWSGPAQCIIIPNK 75  
DB 62 GWSGPAQCIIIPNK 76

RESULT 8  
US-08-769-967A-29  
Sequence 29, Application US/08769967A  
Patent No. 5859223  
GENERAL INFORMATION:  
APPLICANT: Mossakowska, Danuta E.I.  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman, Anne Mary  
TITLE OF INVENTION: Soluble CRI Derivatives  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,967A  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,569  
FILING DATE: 15-May-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: P30423C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5364  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-769-967A-29

Query Match 100.0%; Score 422; DB 2; Length 254;  
Best Local Similarity 100.0%; Pred. No. 5.1e-43;  
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RIPGCLPPTINGDFISTNRENFHYGSVVYTRCNPBGSGGRKVFELVGPSTICTSNDQV 60  
DB 123 RIPGCLPPTINGDFISTNRENFHYGSVVYTRCNPBGSGGRKVFELVGPSTICTSNDQV 182  
QY 61 GWSGPAQCIIIPNK 75  
DB 183 GWSGPAQCIIIPNK 197

RESULT 9  
US-08-769-967A-30  
Sequence 30, Application US/08769967A  
Patent No. 5859223  
GENERAL INFORMATION:  
APPLICANT: Mossakowska, Danuta E.I.  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman, Anne Mary  
TITLE OF INVENTION: Soluble CRI Derivatives  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,967A  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,569  
FILING DATE: 15-May-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: P30423C2  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (610) 270-5364  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-769-967A-30

Query Match 100.0%; Score 422; DB 2; Length 254;  
Best Local Similarity 100.0%; Pred. No. 5,1e-43;  
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPCGLPPTTNDFTSTNENHYGSSVYTYRCNPGSGGKRVELVGEPSICTSDDOY 60  
DB 123 RPPCGLPPTTNDFTSTNENHYGSSVYTYRCNPGSGGKRVELVGEPSICTSDDOY 182  
QY 61 GWSGAPQCIIPNK 75  
DB 183 GWSGAPQCIIPNK 197

## RESULT 10

US-08-769-967A-31  
Sequence 31, Application US/08769967A  
Patent No. 5859223  
GENERAL INFORMATION:  
APPLICANT: Mossakowska, Danuta E.I.  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman, Anne Mary  
TITLE OF INVENTION: Soluble CRI Derivatives  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Smithline Beecham Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,967A  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,569  
FILING DATE: 15-May-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: King William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: P30423C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5364  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 133 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: Internal  
US-08-769-967A-31

Query Match 100.0%; Score 422; DB 2; Length 133;

Best Local Similarity 100.0%; Pred. No. 2.3e-43;  
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPCGLPPTTNDFTSTNENHYGSSVYTYRCNPGSGGKRVELVGEPSICTSDDOY 60  
DB 2 RPPCGLPPTTNDFTSTNENHYGSSVYTYRCNPGSGGKRVELVGEPSICTSDDOY 61  
QY 61 GWSGAPQCIIPNK 75  
DB 62 GWSGAPQCIIPNK 76

## RESULT 11

US-08-310-416A-14  
Sequence 14, Application US/08310416A  
Patent No. 5679546  
GENERAL INFORMATION:  
APPLICANT: Jone-Long Ko et al.  
TITLE OF INVENTION: CHIMERIC PROTEINS WHICH BLOCK  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 555x  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/310,416A  
FILING DATE: 22-SEP-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Paul T. Clark  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 06180/005001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 324 amino acids  
TYPE: amino acid  
STRANDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-310-416A-14

Query Match 37.7%; Score 159; DB 1; Length 324;  
Best Local Similarity 47.8%; Pred. No. 1.4e-11;  
Matches 33; Conservative 8; Mismatches 20; Indels 8; Gaps 3;

QY 2 RPPCGLPPTTNDFTSTNENHYGSSVYTYRCNPGSGGKRVELVGEPSICTSDDOY 61  
DB 189 RPPCGLPPTTNDFTSTNENHYGSSVYTYRCNPGSGGKRVELVGEPSICTSDDOY 240  
QY 62 GWSGAPQCIIPNK 70  
DB 241 GWSGAPQCIIPNK 249

## RESULT 12

US-08-888-171-14  
Sequence 14, Application US/08888171  
Patent No. 5851528  
GENERAL INFORMATION:

APPLICANT: Jone-Long, KO  
APPLICANT: Higgins, Paul J.  
APPLICANT: Yeh, C. Grace  
TITLE OF INVENTION: METHODS OF INHIBITING COMPLEMENT  
TITLE OF INVENTION: ACTIVATION  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson, P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/888,171  
FILING DATE: 03-JUL-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/310,416  
FILING DATE: 22-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Freeman, John W.  
REGISTRATION NUMBER: 29,066  
REFERENCE/DOCKET NUMBER: 06180/005002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-507  
TELEFAX: 617/542-890  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 324 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-888-171-14

Query Match 37.7%; Score 159; DB 2; Length 324;  
Best Local Similarity 47.8%; Pred. No. 1.4e-11;  
Matches 33; Conservative 8; Mismatches 20; Indels 8; Gaps 3;  
QY 2 IPCGLPTITNGDFTSTRENHFGSVVYTYRCNPGSGRGKVEFVGEPSICTSNDQVG 61  
DB 189 IYCPAPQIDNG-IIOGERDHYGRQSVTYACNKG-----FTMIGHSITCYTNNDG 240  
QY 62 IWSGPAQC 70  
DB 241 EMSGPPEPC 249

RESULT 13  
US-08-435-149-2  
Sequence 2, Application US/08435149  
Patent No. 5866402  
GENERAL INFORMATION:  
APPLICANT: INNIS, MICHAEL A.  
APPLICANT: ZAROR, ISABEL  
APPLICANT: CREASEY, ABLA A.  
TITLE OF INVENTION: CHIMERIC MCP AND DAF PROTEINS WITH CELL  
TITLE OF INVENTION: SURFACE LOCALIZING DOMAIN  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CHIRON CORPORATION  
STREET: INTELLECTUAL PROPERTY - R440, P.O. BOX 8097  
CITY: EMERYVILLE  
STATE: CALIFORNIA  
COUNTRY: U.S.A.  
ZIP: 94662-8097  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/435,149  
FILING DATE: 05-MAY-1995  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: SAVEREIDE, PAUL B.  
REGISTRATION NUMBER: 36,914  
REFERENCE/DOCKET NUMBER: 0989.001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-2585  
TELEFAX: (510) 655-3542  
TELEX: N/A  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 323 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-435-149-2

Query Match 37.7%; Score 159; DB 2; Length 323;  
Best Local Similarity 47.8%; Pred. No. 1.4e-11;  
Matches 33; Conservative 8; Mismatches 20; Indels 8; Gaps 3;  
QY 2 IPCGLPTITNGDFTSTRENHFGSVVYTYRCNPGSGRGKVEFVGEPSICTSNDQVG 61  
DB 189 IYCPAPQIDNG-IIOGERDHYGRQSVTYACNKG-----FTMIGHSITCYTNNDG 240  
QY 62 IWSGPAQC 70  
DB 241 EMSGPPEPC 249

RESULT 14  
US-08-435-149-3  
Sequence 3, Application US/08435149  
Patent No. 5866402  
GENERAL INFORMATION:  
APPLICANT: INNIS, MICHAEL A.  
APPLICANT: ZAROR, ISABEL  
APPLICANT: CREASEY, ABLA A.  
TITLE OF INVENTION: CHIMERIC MCP AND DAF PROTEINS WITH CELL  
TITLE OF INVENTION: SURFACE LOCALIZING DOMAIN  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CHIRON CORPORATION  
STREET: INTELLECTUAL PROPERTY - R440, P.O. BOX 8097  
CITY: EMERYVILLE  
STATE: CALIFORNIA  
COUNTRY: U.S.A.  
ZIP: 94662-8097  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/435,149  
FILING DATE: 05-MAY-1995  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: SAVEREIDE, PAUL B.  
REGISTRATION NUMBER: 36,914  
REFERENCE/DOCKET NUMBER: 0989.001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-2585  
TELEFAX: (510) 655-3542







GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 06:07:03 ; Search time 44.47 Seconds

(without alignments)  
63.178 Million cell updates/sec

Title: US-09-142-043-3

Perfect score: 422  
Sequence: 1 RIPCGLPRTTNGDFISTNR.....NDQVGIMSGAPQCIIIPNK 75

Scoring table: BLOSUM62

Searched: 116738 seqs, 37460341 residues

Database: PIR-58:\*\*

1: pir1:\*\*  
2: pir2:\*\*  
3: pir3:\*\*  
4: pir4:\*\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	422	100.0	2014	2	I36936	complement recepto
2	422	100.0	661	2	I36937	complement recepto
3	422	100.0	2489	2	I73012	complement C3b/C4b
4	422	100.0	2039	2	A28507	complement C3b/C4b
5	351	83.2	482	2	A34924	complement C3b/C4b
6	292	69.2	497	2	JC2054	complement regulat
7	291	69.0	433	2	A30550	complement C3b/C4b
8	291	69.0	440	2	A43519	complement recepto
9	254	60.2	676	2	A45900	complement C3d rec
10	193	45.7	1091	1	PL0009	complement C3d/Eps
11	176	41.7	1025	2	A43526	complement C3d/Eps
12	160	37.9	340	2	I56234	decay-accelerating
13	159	37.7	381	1	B26359	decay-accelerating
14	159	37.7	440	2	A26359	decay-accelerating
15	143	33.9	597	2	S53711	C4BP alpha chain p
16	143	33.9	558	2	S57953	C4BP protein alpha
17	132.5	31.4	369	2	I57998	membrane cofactor
18	132.5	31.4	384	2	S01896	membrane cofactor
19	132.5	31.4	377	2	I54479	membrane cofactor
20	132.5	31.4	349	2	G02913	membrane cofactor
21	128.5	30.5	369	2	JC5138	membrane cofactor
22	128.5	30.5	362	2	JC5194	membrane cofactor
23	127.5	30.2	469	1	NBMSCA	C4b-binding protel
24	124.5	29.5	597	1	NBHC4	C4b-binding protel
25	124	29.4	610	2	I46001	C4b-binding protel
26	121	28.7	579	2	A56740	C4b-binding protel
27	114	27.0	668	2	A46013	coagulation factor
28	111.5	26.4	263	2	C36838	coagulation factor
29	109.5	25.9	808	2	D35069	complement control
30	107	25.4	330	2	I55975	complement factor
31	105.5	25.0	1234	1	NBMSH	X/Y protein - mous
32	104.5	24.8	263	1	MWVSP	complement factor
33	100	23.7	363	2	B45900	apolipoprotein H h
34	99	23.5	270	2	I37278	complement C3d/Eps
35	96	22.7	343	2	C35070	complement factor
36	94	22.3	1231	1	NBHUH	apolipoprotein H-r
37	94	22.3	452	2	A35068	complement factor
38	94	22.3	327	2	A40455	complement factor
39	94	22.3	331	2	A45222	complement factor

40	94	22.3	330	2	I56100	factor H homologue
41	94	22.3	330	2	I72653	factor H homologue
42	93	22.0	302	1	WMBE1E	secretory compleme
43	93	22.0	360	1	WMBE2E	membrane-bound com
44	93	22.0	303	2	H35068	apolipoprotein H-r
45	92.5	21.9	449	1	NBHUHS	complement factor

## ALIGNMENTS

## RESULT 1

I36936

complement receptor 1 - chimpanzee (fragment)

C:Species: Pan troglodytes (chimpanzee)

C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 07-Aug-1998

C:Accession: I36936

R:Birmingham, D.J.; Shen, X.P.; Hourcade, D.; Nickells, M.W.; Atkinson, J.P.

J. Immunol. 153, 691-700, 1994

A:Title: Primary sequence of an alternatively spliced form of CRL. Candidate for the

A:Reference number: I36935; MUID:94292799

A:Accession: I36936

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2014 <RES>

A:Cross-references: GB:I24920; NID:9551564; PID:9557725

C:Superfamily: complement factor H repeat homology <FH01>

F:18-74/Domain: complement factor H repeat homology <FH02>

F:79-136/Domain: complement factor H repeat homology <FH03>

F:141-207/Domain: complement factor H repeat homology <FH04>

F:333-391/Domain: complement factor H repeat homology <FH12>

F:1041-1107/Domain: complement factor H repeat homology <FH26>

F:1749-1815/Domain: complement factor H repeat homology <FH27>

## Query Match

Best Local Similarity 100.0%; Score 422; DB 2; Length 2014;

Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## QY

1 RIPCGLPRTTNGDFISTNRNFHYSGVYTRCNPSSGGRKVELGEPISYTSDDQV 60

|||||

Db 138 RIPCGLPRTTNGDFISTNRNFHYSGVYTRCNPSSGGRKVELGEPISYTSDDQV 197

QY 61 GIMSGAPQCIIIPNK 75

|||||

Db 198 GIMSGAPQCIIIPNK 212

## RESULT 2

I36937

complement receptor 1 - chimpanzee (fragment)

C:Species: Pan troglodytes (chimpanzee)

C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 07-Aug-1998

C:Accession: I36937

R:Birmingham, D.J.; Shen, X.P.; Hourcade, D.; Nickells, M.W.; Atkinson, J.P.

J. Immunol. 153, 691-700, 1994

A:Title: Primary sequence of an alternatively spliced form of CRL. Candidate for the

A:Reference number: I36935; MUID:94292799

A:Accession: I36937

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-661 <RES>

A:Cross-references: GB:I24921; NID:9557726; PID:9557727

C:Superfamily: complement factor H repeat homology

F:18-74/Domain: complement factor H repeat homology <FH01>

F:79-136/Domain: complement factor H repeat homology <FH02>

F:141-207/Domain: complement factor H repeat homology <FH03>

F:213-268/Domain: complement factor H repeat homology <FH04>

F:333-391/Domain: complement factor H repeat homology <FH12>

## Query Match

100.0%; Score 422; DB 2; Length 661;

```

OY      1  RIPGCLPTITNGDIFISTIRENFHGYSVYTRCNFGSGRKVFELVGEPSITYCISNDQV 60
          |||||||
Db      138  RIPCLPLPTITNGDIFISTIRENFHGYSVYTRCNFGSGRKVFELVGEPSITYCISNDQV 197
          |||||||
OY      61  GIWGPAPQCIIIPNK 75
          |||||||
Db      198  GIWGPAPQCIIIPNK 212
          |||||||

RESULT  3
173012
complement C3b/C4b receptor (allotype S) precursor - human
N:Alternate names: complement receptor type 1 (CR1); surface glycoprotein CD35
C:Species: Homo sapiens (man)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 24-Sep-1998
C:Accession: I73012; A47602; S03291
R:ViK, D.P.; Mong, W.W.
J: Immunol. 151, 6214-6224, 1993
A:Title: Structure of the gene for the F allele of complement receptor type 1 and sequence
A:Reference number: 156203; MUID:94065175
A:Accession: I73012
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2489 <RES>
A:Cross-references: GB:LI7418; NID:g306678; PID:g451303
R:Mong, W.W.; Cahill, J.M.; Rosen, M.D.; Kennedy, C.A.; Bonaccio, E.T.; Morris, M.J.; W
J: Exp. Med. 169, 847-863, 1989
A:Title: Structure of the human CR1 gene. Molecular basis of the structural and quantita
A:Reference number: A47602
A:Accession: A47602
A:Molecule type: DNA
A:Residues: 1-41 <MON>
A:Cross-references: GB:X14893
J: Hourcade, D.; Miesner, D.R.; Atkinson, J.P.; Holers, V.M.
A:Title: Identification of an alternative polyadenylation site in the human C3b/C4b rece
type 1.
A:Reference number: S03291; MUID:89010527
A:Accession: S03291
A:Molecule type: mRNA
A:Residues: 26-584 <HOU>
A:Cross-references: EMBL:X14362; NID:g30197; PID:g736240
A:Experimental source: clone CR1-4
C:Genetics:
A:Gene: GDB:CR1; CD35
A:Cross-references: GDB:119800; OMIM:120620
A:Map position: 1q32-1q32
A:Intons: 41/1; 101/1; 134/2; 163/1; 296/1; 355/1; 389/2; 418/1; 491/1; 551/1; 584/2; 6
; 1444/2; 1513/1; 1646/1; 1705/1; 1739/2; 1768/1; 1844/1; 1904/1; 1937/2; 1966/1; 2099/2
C:Superfamily: complement factor H repeat homology
C:Keywords: duplication; glycoprotein; transmembrane protein
F:1-41/Domain: signal sequence #status predicted <STR>
F:42-584/Product: complement C3b/C4b receptor, secreted #status predicted <MAN>
F:43-99/Domain: complement factor H repeat homology <FHO1>
F:104-161/Domain: complement factor H repeat homology <FHO2>
F:166-232/Domain: complement factor H repeat homology <FHO3>
F:238-293/Domain: complement factor H repeat homology <FHO4>
F:297-353/Domain: complement factor H repeat homology <FHO5>
F:358-416/Domain: complement factor H repeat homology <FHO6>
F:421-487/Domain: complement factor H repeat homology <FHO7>
F:554-611/Domain: complement factor H repeat homology <FHO8>
F:616-682/Domain: complement factor H repeat homology <FHO9>
F:808-866/Domain: complement factor H repeat homology <FHO10>
F:1004-1061/Domain: complement factor H repeat homology <FHO11>
F:1066-1132/Domain: complement factor H repeat homology <FHO12>
F:1138-1193/Domain: complement factor H repeat homology <FHO13>
F:1197-1253/Domain: complement factor H repeat homology <FHO14>
F:1258-1316/Domain: complement factor H repeat homology <FHO15>
F:1321-1387/Domain: complement factor H repeat homology <FHO16>
F:1393-1449/Domain: complement factor H repeat homology <FHO17>

```

```

F:1454-1511/Domain: complement factor H repeat homology <FH1>
F:1516-1582/Domain: complement factor H repeat homology <FH1>
F:1588-1643/Domain: complement factor H repeat homology <FH1>
F:1647-1703/Domain: complement factor H repeat homology <FH1>
F:1708-1766/Domain: complement factor H repeat homology <FH2>
F:1771-1837/Domain: complement factor H repeat homology <FH2>
F:1846-1902/Domain: complement factor H repeat homology <FH2>
F:1907-1964/Domain: complement factor H repeat homology <FH2>
F:1969-2035/Domain: complement factor H repeat homology <FH2>
F:2100-2156/Domain: complement factor H repeat homology <FH2>
F:2161-2219/Domain: complement factor H repeat homology <FH2>

Query Match      100.0%; Score 422; DB 2; Length 2489;
Best Local Similarity 100.0%; Pred. No. 1,1e-38;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1  RIPGLPPTINGEISINRENFHYSGVYVYRCNPGSGGRVFEELPEPSITCTSNDDV 60
Db      163 RIPGLPPTINGDFISINRENFHYSGVYVYTRCNPGSGGRVFEELPEPSITCTSNDDV 222

Oy      61 GIMSGPAPQCIIIPNK 75
Db      223 GIMSGPAPQCIIIPNK 237

RESULT      4
A28507
Complement C3b/C4b receptor precursor, membrane-bound (allotype F) - human
N:Alternate names: complement receptor type 1 (CR1); surface glycoprotein CD35
C:Species: Homo sapiens (man)
C:Date: 19-Nov-1998 #sequence_revision 05-Sep-1996 #text_change 10-Sep-1997
C:Accession: S03843; A28507; I56203; A24748; B24748; C24748
R:Klickstein, L.B.; Barlow, T.J.; Miletic, V.; Rabson, L.D.; Smith, J.A.; Fearon, D.T.
J. Exp. Med. 168, 1699-1717, 1988
A:Title: Identification of distinct C3b and C4b recognition sites in the human C3b/C4
A:Reference number: S03843; MUID:89035992
A:Accession: S03843
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-2039 <KLI>
A:Cross-references: EMBL:Y00816; NID:G30186
R:Klickstein, L.B.; Wong, W.W.; Smith, J.A.; Wels, J.H.; Wilson, J.G.; Fearon, D.T.
J. Exp. Med. 165, 1095-1112, 1987
A:Title: Human C3b/C4b receptor (CR1). Demonstration of long homologous repeating dom
A:Reference number: A28507; MUID:87168191
A:Accession: A28507
A:Molecule type: mRNA
A:Residues: 503-771,'FV',774-2039 <KLT>
R:VLK, D.P.; Wong, W.W.
J. Immunol. 151, 6214-6224, 1993
A:Title: Structure of the gene for the F allele of complement receptor type 1 and seq
A:Reference number: I56203; MUID:94065175
A:Accession: I56203
A:Status: translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-683,'X',685-1021,'X',1023-1614,'V',1616-1836,'R',1828-1849,'D',1851-1872
A:Cross-references: GB:U7418; NID:G306878; PID:G306860
R:Mong, W.W.; Klickstein, L.B.; Smith, J.A.; Wels, J.H.; Fearon, D.T.
Proc. Natl. Acad. Sci. U.S.A. 82, 7711-7715, 1985
A:Title: Identification of a partial cDNA clone for the human receptor for complement
A:Reference number: A94073; MUID:8606975
A:Accession: A24748
A:Molecule type: mRNA
A:Residues: 311-333;729-745;831-845 <KON>
C:Genetics:
A:Gene: GDB:CR1; CD35
A:Cross-references: GDB:119800; OMIM:120620
A:Map position: 1932-1932
A:Introns: 41/1; 101/1; 134/2; 163/1; 296/1; 355/1; 389/2; 418/1; 491/1; 551/1; 584/2
1487/2; 1516/1; 1649/1; 1708/1; 1742/2; 1771/1; 1841/1; 1906/1; 1968/1; 1976/1; 200
C:Superfamily: complement factor H repeat homology
C:Keywords: duplication; glycoprotein; transmembrane protein

```

F:1-41/Domain: signal sequence #status predicted <SIG>  
F:12-2039/Product: complement C3b/C4b receptor, membrane-bound #status predicted <MAT>  
F:104-161/Domain: complement factor H repeat homology <FH01>  
F:104-161/Domain: complement factor H repeat homology <FH02>  
F:166-232/Domain: complement factor H repeat homology <FH03>  
F:238-293/Domain: complement factor H repeat homology <FH04>  
F:297-353/Domain: complement factor H repeat homology <FH05>  
F:358-416/Domain: complement factor H repeat homology <FH06>  
F:421-487/Domain: complement factor H repeat homology <FH07>  
F:493-549/Domain: complement factor H repeat homology <FH08>  
F:554-611/Domain: complement factor H repeat homology <FH09>  
F:616-682/Domain: complement factor H repeat homology <FH10>  
F:688-743/Domain: complement factor H repeat homology <FH11>  
F:747-803/Domain: complement factor H repeat homology <FH12>  
F:808-866/Domain: complement factor H repeat homology <FH13>  
F:871-937/Domain: complement factor H repeat homology <FH14>  
F:943-999/Domain: complement factor H repeat homology <FH15>  
F:1004-1061/Domain: complement factor H repeat homology <FH16>  
F:1066-1132/Domain: complement factor H repeat homology <FH17>  
F:1138-1193/Domain: complement factor H repeat homology <FH18>  
F:1197-1253/Domain: complement factor H repeat homology <FH19>  
F:1258-1316/Domain: complement factor H repeat homology <FH20>  
F:1321-1387/Domain: complement factor H repeat homology <FH21>  
F:1396-1452/Domain: complement factor H repeat homology <FH22>  
F:1457-1514/Domain: complement factor H repeat homology <FH23>  
F:1519-1585/Domain: complement factor H repeat homology <FH24>  
F:1591-1646/Domain: complement factor H repeat homology <FH25>  
F:1650-1706/Domain: complement factor H repeat homology <FH26>  
F:1711-1769/Domain: complement factor H repeat homology <FH27>  
F:1774-1840/Domain: complement factor H repeat homology <FH28>  
F:1848-1904/Domain: complement factor H repeat homology <FH29>  
F:1909-1965/Domain: complement factor H repeat homology <FH30>

Query Match 100.0% Score 422; DB 2; Length 2039;  
Best Local Similarity 100.0% Pred. No. 8.4e-39;  
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RIFCGLPPTTNDGFISTNRNFHYGSVVYTRCNPSSGGRKVELGEPISYCTSDDDV 60  
|||||  
Db 163 RIFCGLPPTTNDGFISTNRNFHYGSVVYTRCNPSSGGRKVELGEPISYCTSDDDV 222

QY 61 GWSGAPQCIIPNK 75  
|||||  
Db 223 GWSGAPQCIIPNK 237

RESULT 5  
A34924  
complement C3b/C4b receptor-like protein precursor - human  
C:Species: Homo sapiens (man)  
C:Date: 27-Jul-1990 #sequence\_revision 11-Apr-1997 #text\_change 11-Apr-1997  
C:Accession: A34924; S03292  
R:Hourcade, D.; Miesner, D.R.; Bee, C.; Zeides, W.; Atkinson, J.P.  
J. Biol. Chem. 265, 974-980, 1990  
A:Title: Duplication and divergence of the amino-terminal coding region of the complemen  
A:Reference number: A34924; MUID:90110163  
A:Accession: A34924  
A:Molecule type: DNA  
A:Residues: 1-479 <H00>  
A:Cross-references: GB:J05195  
R:Hourcade, D.; Miesner, D.R.; Atkinson, J.P.; Holers, V.M.  
J. Exp. Med. 168, 1255-1270, 1988  
A:Title: Identification of an alternative polyadenylation site in the human C3b/C4b rece  
type 1.  
A:Reference number: S03291; MUID:89010527  
A:Accession: S03292  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 'PPPPFAFR', 452-482 <H02>  
A:Cross-references: EMBL:X14360  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1988  
C:Superfamily: complement factor H repeat homology

F:43-99/Domain: complement factor H repeat homology <FH01>  
F:104-161/Domain: complement factor H repeat homology <FH02>  
F:166-232/Domain: complement factor H repeat homology <FH03>  
F:238-293/Domain: complement factor H repeat homology <FH04>  
F:297-353/Domain: complement factor H repeat homology <FH05>  
F:358-416/Domain: complement factor H repeat homology <FH06>  
F:421-478/Domain: complement factor H repeat homology <FH07>

Query Match 83.2% Score 351; DB 2; Length 482;  
Best Local Similarity 85.3% Pred. No. 1.3e-31;  
Matches 64; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 RIFCGLPPTTNDGFISTNRNFHYGSVVYTRCNPSSGGRKVELGEPISYCTSDDDV 60  
|||||  
Db 163 RIFCGLPPTTNDGFISTNRNFHYGSVVYTRCNPSSGGRKVELGEPISYCTSDDDV 222

QY 61 GWSGAPQCIIPNK 75  
|||||  
Db 223 GWSGAPQCIIPNK 237

RESULT 6  
JC2054  
complement regulatory protein, 512 antigen precursor - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 27-Jun-1994 #sequence\_revision 27-Jun-1994 #text\_change 22-Nov-1996  
C:Accession: JC2054; PC2027  
R:Sakurada, C.; Seno, H.; Dohl, N.; Takizawa, H.; Nonaka, M.; Okada, N.; Okada, H.  
Biochem. Biophys. Res. Commun. 198, 819-826, 1994  
A:Title: Molecular cloning of the rat complement regulatory protein, 512 antigen.  
A:Reference number: JC2054  
A:Accession: JC2054  
A:Molecule type: mRNA  
A:Residues: 1-497 <SAK>  
A:Accession: PC2027  
A:Molecule type: protein  
A:Residues: 39-51 <SAZ>  
C:Comment: This protein plays a critical role in protection against complement mediat  
C:Superfamily: complement factor H repeat homology  
C:Keywords: glycoprotein; transmembrane protein  
F:1-36/Domain: signal sequence #status predicted <SIG>  
F:37-488/Product: complement regulatory protein, 512 antigen #status predicted <MAT>  
F:38-94/Domain: complement factor H repeat homology <FH1>  
F:99-156/Domain: complement factor H repeat homology <FH2>  
F:161-227/Domain: complement factor H repeat homology <FH3>  
F:233-286/Domain: complement factor H repeat homology <FH4>  
F:294-352/Domain: complement factor H repeat homology <FH5>  
F:357-413/Domain: complement factor H repeat homology <FH6>  
F:421-444/Domain: transmembrane #status predicted <TM>  
F:247,331,346,450,482,483/Binding site: carbohydrate (asn) (covalent) #status predict

Query Match 69.2% Score 292; DB 2; Length 497;  
Best Local Similarity 68.9% Pred. No. 5e-25;  
Matches 51; Conservative 9; Mismatches 14; Indels 0; Gaps 0;

QY 2 IFCGLPPTTNDGFISTNRNFHYGSVVYTRCNPSSGGRKVELGEPISYCTSDDDV 61  
|||||  
Db 159 IFCEIPPSIPNDGFPNEDFHYGMVYVQCNTDARGKRLNLVGEPSIHCTSIDGVC 218

QY 62 IWSGAPQCIIPNK 75  
|||||  
Db 219 IWSGAPQCIIPNK 232

RESULT 7  
A30550  
complement C3b/C4b receptor precursor - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 03-Mar-1989 #sequence\_revision 03-Mar-1989 #text\_change 12-May-1995  
C:Accession: A30550  
R:Paul, M.S.; Aegerter, M.; O'Brien, S.E.; Kurtz, C.B.; Weis, J.H.

A>Title: The murine complement receptor gene family. IV. Alternative splicing of Cr2  
A:Reference number: A45900  
A:Accession: A45900  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-676 <KUR>  
A:Cross-references: GB:M36470  
A:Experimental source: GB:M36470  
R:Kim, Y.U.; Kinoshita, T.; Molina, H.; Hourcade, D.; Seya, T.; Wagner, L.M.; Holters,  
J. Exp. Med. 181, 151-159, 1995  
A>Title: Mouse complement regulatory protein Crry/p65 uses the specific mechanisms of  
A:Reference number: I48306; MUID:95105691  
A:Accession: I48306  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 21-367 <RES>  
A:Cross-references: EMBL:U17128; NID:g9595980; PID:g9595982  
C:Genetics:  
A:Gene: Cr2  
A:Introns: 80/1; 113/2; 142/1; 275/1; 333/1  
C:Superfamily: complement factor H repeat homology  
F:2-278/Domain: complement factor H repeat homology <FH01>  
F:83-140/Domain: complement factor H repeat homology <FH02>  
F:145-211/Domain: complement factor H repeat homology <FH03>  
F:217-272/Domain: complement factor H repeat homology <FH04>  
F:276-331/Domain: complement factor H repeat homology <FH05>  
F:336-394/Domain: complement factor H repeat homology <FH06>  
F:399-458/Domain: complement factor H repeat homology <FH07>  
F:467-523/Domain: complement factor H repeat homology <FH08>  
F:531-587/Domain: complement factor H repeat homology <FH09>  
F:592-648/Domain: complement factor H repeat homology <FH10>

Query Match            60.2%; Score 254; DB 2; Length 676;  
Best Local Similarity   62.9%; Pred. No. 1.2e-20;  
Matches     44; Conservative   10; Mismatches   16; Indels     0; Gaps     0;

OY       2   IPCGLPTTNGDFTSTRENFHYGSAVYYKCNPGSGRKVFELVGEPSIYCSTNDQVG   61  
         |||   ||   |||||   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :  
Db      143   IPCSPALISNGDDSSSRDSFFFGMYVTYYCHCKNKRKEFLDLVGESIKYCTSKDNQVG   202

OY       62   IMSGAPQCI   71  
         ||:   |   |||||  
Db      203   IMNSPPQCI   212

RESULT   10  
PL0009  
complement C3d/Epstein-Barr virus receptor precursor - human  
N:Alternate names: complement receptor 2; CR2/CD21  
C:Species: Homo sapiens (man)  
C>Date: 30-Jun-1992 \$sequence\_revision 07-Jul-1995 #text\_change 05-Sep-1997  
C:Accession: J10028; A33958; A32036; A24319; B24319; C24319; D24319; E24319; F24319;  
R:Weis, J.J.; Tothaker, L.E.; Smith, J.A.; Weis, J.H.; Fearon, D.T.  
A:Exp. Med. 167, 1047-1066, 1988  
A>Title: Structure of the human B lymphocyte receptor for C3d and the Epstein-Barr viru  
A:Reference number: J10028; MUID:81171282  
A:Accession: J10028  
A:Molecule type: mRNA  
A:Residues: 1-1091 <WEI>  
A>Note: nucleotides 1566-1625 are missing from Figure 1; therefore, residues 522-542  
R:Moore, M.D.; Cooper, N.R.; Tack, B.F.; Nemereow, G.R.  
Proc. Natl. Acad. Sci. U.S.A. 84, 9194-9198, 1987  
A>Title: Molecular cloning of the cDNA encoding the Epstein-Barr virus/C3d receptor (  
A:Reference number: A33958; MUID:88097454  
A:Accession: A33958  
A:Molecule type: mRNA  
A:Residues: 1-456; 'G', 457-644, 'R', 646-669, 'R', 671-816, 'NCSEVILKWLIERAF', 835-840, 'L', 'H'  
A:Cross-references: GB:I03365; NID:g181919; PID:g181920  
R:Pujols, A.; Harley, J.B.; Frank, M.B.; Gruner, B.A.; Frazier, B.; Holters, V.M.  
Biochim. Chem. 264, 2118-2125, 1989  
A>Title: Genomic organization and polymorphisms of the human C3d/Epstein-Barr virus r  
A:Reference number: A32036; MUID:89132277



A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-340 <RES>  
A:Cross-references: GB:S67775; NID:q459389; PID:q459390  
C:Superfamily: decay-accelerating factor; complement factor H repeat homology  
F:1-53/Domain: complement factor H repeat homology (fragment) <FH01>  
F:122-179/Domain: complement factor H repeat homology <FH02>  
F:184-242/Domain: complement factor H repeat homology <FH03>

Query Match 37.9%; Score 160; DB 2; Length 340;  
Best Local Similarity 46.4%; Pred. No. 1,6e-10;  
Matches 32; Conservative 9; Mismatches 20; Indels 8; Gaps 3;

QY 2 IPCGLPPTTNGDFTSTNRNENHGVVYTRCNPQSGRKVFELVGPSTICTSNDQVG 61  
DB 182 IYCAPPQIDNG-IIGCRDHRHGYRQSTVYACNKG-----YTMIGHSTICTVNDDE-G 233  
QY 62 IWSGPAPQC 70  
DB 234 EMSGPPPEC 242

RESULT 13  
B26359  
N:Alternate names: CD55; DAF splice form 2; decay-accelerating factor membrane-bound for C:Species: Homo sapiens (mn)  
C:Date: 05-Oct-1988 #sequence,revision 16-Aug-1996 #text,change 24-Apr-1998  
C:Accession: B26359; A27666; A39101; 152594; 152564  
R:Caras, I.W.; Daviltz, M.A.; Rhee, L.; Weddell, G.; Martin Jr., D.W.; Nussenzweig, V.  
Nature 325, 545-549, 1987  
A:Title: Cloning of decay-accelerating factor suggests novel use of splicing to generate  
A:Reference number: A26359; MUID:87115845  
A:Accession: B26359  
A:Molecule type: mRNA  
A:Residues: 1-381 <CAR>  
A:Cross-references: GB:M30142; NID:g181464; PID:g181465  
R:Medof, M.E.; Lublin, D.M.; Holers, V.M.; Ayers, D.J.; Getty, R.R.; Laykam, J.F.; Atkin  
Proc. Natl. Acad. Sci. U.S.A. 84, 2007-2011, 1987  
A:Title: Cloning and characterization of cDNAs encoding the complete sequence of decay-a  
A:Reference number: A27666; MUID:87175602  
A:Accession: A27666  
A:Molecule type: mRNA  
A:Residues: 6-79, 'T', '81-84', 'M', '86-381 <MED>  
A:Cross-references: GB:M15799; NID:g181462; PID:g181463  
R:Moran, P.; Raab, H.; Kohr, W.J.; Caras, I.W.  
J. Biol. Chem. 266, 1250-1257, 1991  
A:Title: Glycophospholipid membrane anchor attachment. Molecular analysis of the cleavag  
A:Reference number: A39101; MUID:91093238  
A:Accession: A39101  
A:Molecule type: protein  
A:Residues: 338-352 <MOR>  
R:Lublin, D.M.; Mallinson, G.; Poole, J.; Reid, M.E.; Thompson, E.S.; Ferdman, B.R.; Tel  
Blood 84, 1276-1282, 1994  
A:Title: Molecular basis of reduced or absent expression of decay-accelerating factor in  
A:Reference number: 152594  
A:Accession: 152594  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 194-198, 'L', 200-209 <LUB>  
A:Cross-references: GB:S72858; NID:g639599; PID:g639600  
A:Experimental source: individual KW, Cromer blood group phenotype Dr(a-)  
A:Note: the single nucleotide difference in this allele, which changes Ser-199 to Leu, 1  
on (see reference 152564), and thus reduced DAF expression  
R:Reid, M.E.; Mallinson, G.; SIm, R.B.; Poole, J.; Pausch, V.; Merry, A.H.; Ilew, Y.W.;  
Blood 78, 3291-3297, 1991  
A:Title: Biochemical studies on red blood cells from a patient with the Inab phenotype  
A:Reference number: 152564; MUID:92075980  
A:Accession: 152564  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 190-193, 'OLCPVE' <RE2>

A:Cross-references: GB:S70688; NID:q240301; PID:q240302  
A:Experimental source: individual KW, Cromer blood group phenotype Dr(a-) (described  
C:Comment: Cromer blood group system antigens reside on this protein.  
C:Genetics:  
A:Gene: GDB:DAF  
A:Cross-references: GDB:119088; OMIM:125240  
A:Map position: 1q32-1q32  
C:Function:  
A:Description: protects tissues from damage by regulating complement activation on ce  
C:Superfamily: decay-accelerating factor; complement factor H repeat homology  
C:Keywords: alternative splicing; complement inhibitor; glycoprotein; lipoprotein; ph  
F:1-34/Domain: signal sequence #status predicted <SIG>  
F:33-353/Product: decay-accelerating factor 2 #status predicted <MAT>  
F:36-94/Domain: complement factor H repeat homology <FH01>  
F:98-158/Domain: complement factor H repeat homology <FH02>  
F:163-220/Domain: complement factor H repeat homology <FH03>  
F:225-283/Domain: complement factor H repeat homology <FH04>  
F:354-381/Domain: carboxyl-terminal propeptide #status predicted <PRO>  
F:95/Binding site: carboxylate (Asn) (covalent) #status predicted  
F:353/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Ser) (in mature f

Query Match 37.7%; Score 159; DB 1; Length 381;  
Best Local Similarity 47.8%; Pred. No. 2,3e-10;  
Matches 33; Conservative 8; Mismatches 20; Indels 8; Gaps 3;

QY 2 IPCGLPPTTNGDFTSTNRNENHGVVYTRCNPQSGRKVFELVGPSTICTSNDQVG 61  
DB 223 IYCAPPQIDNG-IIGCRDHRHGYRQSTVYACNKG-----FTWIGHSTICTVNDDE-G 274  
QY 62 IWSGPAPQC 70  
DB 275 EMSGPPPEC 283

RESULT 14  
A26359  
N:Alternate names: decay-splice form 1 precursor - human  
C:Species: Homo sapiens (mn)  
C:Date: 05-Oct-1988 #sequence,revision 05-Oct-1988 #text,change 29-Aug-1997  
C:Accession: A26359; A39702; S16187; S23138; A27258  
R:Caras, I.W.; Daviltz, M.A.; Rhee, L.; Weddell, G.; Martin Jr., D.W.; Nussenzweig, V.  
Nature 325, 545-549, 1987  
A:Title: Cloning of decay-accelerating factor suggests novel use of splicing to gener  
A:Reference number: A26359; MUID:87115845  
A:Accession: A26359  
A:Molecule type: mRNA  
A:Residues: 1-440 <CAR>  
A:Cross-references: GB:M30142  
R:Emulonu, U.R.; Ravi, L.; Medof, M.E.  
Proc. Natl. Acad. Sci. U.S.A. 88, 4675-4679, 1991  
A:Title: Characterization of the decay-accelerating factor gene promoter region.  
A:Reference number: A39702; MUID:91271256  
A:Accession: A39702  
A:Molecule type: DNA  
A:Residues: 1-79, 'T', '81-104 <EMU>  
A:Cross-references: GB:M64356  
A:Note: the authors translated the codon AGT for residue 85 as Met  
R:Nakano, Y.; Sugita, Y.; Ishikawa, Y.; Choi, N.H.; Tobe, T.; Tomita, M.  
Biochim. Biophys. Acta 1074, 326-330, 1991  
A:Title: Isolation of two forms of decay-accelerating factor (DAF) from human urine.  
A:Reference number: S16187; MUID:91291869  
A:Accession: S16187  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 35-47 <BIO>  
R:Nakano, Y.; Sunida, K.; Kikuta, N.; Miura, N.H.; Tobe, T.; Tomita, M.  
Biochim. Biophys. Acta 1116, 235-240, 1992  
A:Title: Complete determination of disulfide bonds localized within the short consens  
A:Reference number: S23138  
A:Accession: S23138  
A:Status: preliminary







GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 11:37:50 ; Search time 31.53 seconds

(without alignments)  
63.845 Million cell updates/sec

Title: US-09-142-043-3

Perfect score: 422

Sequence: 1 R1PCGDPPTTNGDPISTNR.....NDQVIGWSPAPQCITPNK 75

Scoring table: BLOSUM62

Searched: 74019 seqs, 26840295 residues

Database: Swissprot\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	422	100.0	2039	1	CRI_HUMAN
2	193	45.7	1033	1	P17927 homo sapien
3	176	41.7	1025	1	P20023 homo sapien
4	160	37.9	340	1	P19070 mus musculu
5	159	37.7	381	1	P49457 pongo pygma
6	148.5	35.2	350	1	P08174 homo sapien
7	143	33.9	558	1	P06147 mus musculu
8	132.5	31.4	377	1	P06154 rattus norv
9	130.5	30.9	407	1	P15529 homo sapien
10	128	30.3	507	1	P06147 mus musculu
11	127.5	30.2	469	1	P06401 cavia porce
12	124.5	29.5	557	1	P08607 mus musculu
13	124	29.4	610	1	P04003 homo sapien
14	114	27.0	668	1	P02805 bos taurus
15	105.5	25.0	1234	1	P07968 mus musculu
16	104.5	24.8	263	1	P06909 mus musculu
17	99	23.5	270	1	P10998 vaccinia vl
18	97	23.0	958	1	P36960 homo sapien
19	94	22.3	1231	1	P09101 drosophila
20	94	22.3	330	1	P08603 homo sapien
21	94	22.3	331	1	P03501 homo sapien
22	93	22.0	360	1	P02985 homo sapien
23	92	21.8	2109	1	P01016 herpesvirus
24	92	21.8	1268	1	P07898 gallus gall
25	92	21.6	2124	1	P55066 mus musculu
26	89	21.1	661	1	P07897 rattus norv
27	89	21.1	2132	1	P05160 homo sapien
28	88	20.9	3562	1	P01282 mus musculu
29	87	20.6	3258	1	P09093 gallus gall
30	87	20.6	3358	1	P06315 rattus norv
31	86	20.4	372	1	P06205 mus musculu
32	86	20.4	372	1	P05198 macaca mula
33	86	20.4	883	1	P028768 papio hamad
34	86	20.4	883	1	P01361 mus musculu
35	86	20.4	1257	1	P55068 rattus norv
36	86	20.4	1257	1	P55067 rattus norv
37	85	19.7	768	1	P13611 homo sapien
38	83	19.7	1019	1	P98106 rattus norv
39	82	19.4	564	1	P28175 tachyples
40	81.5	19.3	758	1	P028343 canis faml
41	81	19.2	372	1	P001102 mus musculu
42	81	19.2	372	1	P14151 homo sapien
43	81	19.2	372	1	P095237 pan troglod
					P095235 pongo pygma

## ALIGNMENTS

RESULT	1	44	81	19.2	933	1	PERT_HUMAN
CRI_HUMAN		45	80	19.0	688	1	CIS_HUMAN
ID	CRI_HUMAN						
AC	P17927						
DT	01-NOV-1990 (REL. 16, CREATED)						
DT	01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)						
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)						
DE	COMPLEMENT RECEPTOR TYPE 1 PRECURSOR (C3B/C4B RECEPTOR) (CD35 ANTIGEN).						
DE	ANTIGEN).						
GN	CRI OR C3BR.						
OS	HOMO SAPIENS (HUMAN).						
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA; EUTHERIA; PRIMATES.						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RX	MEDLINE: 89035992.						
RA	KLICKSTEIN L.B., BARTOW T.J., MILETIC V., RABSON L.D., SMITH J.A., FEARON D.T.;						
RL	J. EXP. MED. 168:1699-1717(1988).						
RN	[2]						
RP	SEQUENCE OF 503-2039 FROM N.A.						
RX	MEDLINE: 87168191.						
RA	KLICKSTEIN L.B., WONG W.W., SMITH J.A., WEIS J.H., WILSON J.G., FEARON D.T.;						
RL	J. EXP. MED. 165:1095-1112(1987).						
RP	SEQUENCE OF 761-783; 831-845 AND 1179-1195 FROM N.A.						
RX	MEDLINE: 86067975.						
RA	WONG W.W., KLICKSTEIN L.B., SMITH J.A., WEIS J.H., FEARON D.T.;						
RL	PROC. NATL. ACAD. SCI. U.S.A. 82:7711-7715(1985).						
CC	-1- FUNCTION: CRI, PRESENT ON ERYTHROCYTES, LEUCOCYTES, GLOMERULAR PODOCYTES, AND SPERMIC FOLLICULAR DENDRITIC CELLS, MEDIATES THE BINDING BY THESE CELLS OF PARTICLES AND IMMUNE COMPLEXES THAT HAVE ACTIVATED COMPLEMENT.						
CC	-1- SUBUNIT: MONOMER.						
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.						
CC	-1- POLYMORPHISM: CRI IS RESPONSIBLE FOR THE KNOX BLOOD GROUP SYSTEM.						
CC	-1- SEVEN SHORT CONSENSUS REPEATS (SCR) CONSTITUTE A LONG HOMOLOGOUS REPEAT (LHR). THE N-TERMINAL TWO SCRS OF LHR-A CONTAINED A SITE DETERMINING C4 SPECIFICITY, AND THE N-TERMINAL TWO SCRS OF LHR-B CONTAINED A SITE DETERMINING C3 SPECIFICITY.						
CC	-1- THIS IS THE SEQUENCE OF THE F ALLOTYPIC OF CRI.						
CC	-1- SIMILARITY: CONTAINS 30 SUSHI (SCR) REPEATS.						
DR	EMBL: M11569; G180995; -						
DR	EMBL: M11617; G180996; -						
DR	EMBL: Y00816; G30186; -						
DR	EMBL: X05309; G809019; -						
DR	PIR: A28507; A28507; -						
DR	PIR: A24748; A24748; -						
DR	PIR: B24748; B24748; -						
DR	PIR: C24748; C24748; -						
DR	PIR: S03843; S03843; -						
DR	HSSP: P08603; LHF1.						
DR	MIM: 120620; -						
KW	COMPLEMENT PATHWAY; GLYCOPROTEIN; TRANSMEMBRANE; REPEAT; SIGNAL; RECEPTOR; SUSHI; BLOOD GROUP ANTIGEN.						
KW	SIGNAL						
FT	CHAIN	1	41				
FT	DOMAIN	42	2039				
FT	DOMAIN	42	1971				
FT	TRANSMEM	1972	1996				
FT	DOMAIN	1997	2039				
FT	MOD_RES	42	42				
FT	DOMAIN	42	488				
FT	REPEAT	42	100				
FT	REPEAT	42	100				
FT	SUSHI A1.						

FT	DISULFID	1226	1253	BY SIMILARITY.	
FT	DISULFID	1258	1300	BY SIMILARITY.	
FT	DISULFID	1286	1316	BY SIMILARITY.	
FT	DISULFID	1321	1370	BY SIMILARITY.	
FT	DISULFID	1350	1387	BY SIMILARITY.	
FT	DISULFID	1396	1439	BY SIMILARITY.	
FT	DISULFID	1426	1452	BY SIMILARITY.	
FT	DISULFID	1457	1498	BY SIMILARITY.	
FT	DISULFID	1484	1514	BY SIMILARITY.	
FT	DISULFID	1519	1568	BY SIMILARITY.	
FT	DISULFID	1548	1585	BY SIMILARITY.	
FT	DISULFID	1591	1633	BY SIMILARITY.	
FT	DISULFID	1619	1646	BY SIMILARITY.	
FT	DISULFID	1650	1693	BY SIMILARITY.	
FT	DISULFID	1679	1706	BY SIMILARITY.	
FT	DISULFID	1711	1753	BY SIMILARITY.	
FT	DISULFID	1739	1769	BY SIMILARITY.	
FT	DISULFID	1774	1823	BY SIMILARITY.	
FT	DISULFID	1803	1840	BY SIMILARITY.	
FT	DISULFID	1848	1891	BY SIMILARITY.	
FT	DISULFID	1877	1904	BY SIMILARITY.	
FT	DISULFID	1909	1952	BY SIMILARITY.	
FT	DISULFID	1938	1965	BY SIMILARITY.	
FT	CARBOHYD	56	56	POTENTIAL.	
FT	CARBOHYD	252	252	POTENTIAL.	
FT	CARBOHYD	410	410	POTENTIAL.	
FT	CARBOHYD	447	447	POTENTIAL.	
FT	CARBOHYD	509	509	POTENTIAL.	
FT	CARBOHYD	578	578	POTENTIAL.	
FT	CARBOHYD	702	702	POTENTIAL.	
FT	CARBOHYD	860	860	POTENTIAL.	
FT	CARBOHYD	897	897	POTENTIAL.	
FT	CARBOHYD	959	959	POTENTIAL.	
FT	CARBOHYD	1028	1028	POTENTIAL.	
FT	CARBOHYD	1152	1152	POTENTIAL.	
FT	CARBOHYD	1310	1310	POTENTIAL.	
FT	CARBOHYD	1481	1481	POTENTIAL.	
FT	CARBOHYD	1504	1504	POTENTIAL.	
FT	CARBOHYD	1534	1534	POTENTIAL.	
FT	CARBOHYD	1540	1540	POTENTIAL.	
FT	CARBOHYD	1605	1605	POTENTIAL.	
FT	CARBOHYD	1763	1763	POTENTIAL.	
FT	CARBOHYD	1908	1908	POTENTIAL.	
SO	SEQUENCE	2039	AA: 223589	MM: 666F9033	CRC32;
Query Match					
Best Local Similarity 100.0%; Score 422; DB 1; Length 2039;					
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps					
OY	1 RIPCGLPPTINGDEFISTNRENFHYSVYVYRCNPGSGGRKVFELVGEPSIYCTSNDOY 60				
Db	163 RIPCGLPPTINGFISITNRENFHYSVYVYRCNPGSGGRKVFELVGEPSIYCTSNDOY 222				
OY	61 GIWSGAPQCIIPNK 75				
Db	223 GIWSGAPQCIIPNK 237				
RESULT 2					
ID	CR2_HUMAN	STANDARD;	PRT;	1033	AA.
AC	P20023;				
DT	01-FEB-1991 (REL. 17, CREATED)				
DT	01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)				
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)				
DE	COMPLEMENT RECEPTOR TYPE 2 PRECURSOR (CR2) (COMPLEMENT C3D RECEPTOR)				
DE	(EBSTEIN-BARR VIRUS RECEPTOR) (CD21 ANTIGEN).				
GN	CR2 OR CD2R				
OS	HOMO SAPIENS (HUMAN).				
OC	EUDAROTIA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;				
OC	EUTHERIA; PRIMATES.				
RN	[1]				

RP SEQUENCE FROM N.A.  
RX MEDLINE: 89123277.  
RA FUJISAKU A., HARLEY J.B., FRANK M.B., GRUNER B.A., FRAZIER B.,  
RA HOLERS V.M.;  
RL J. BIOL. CHEM. 264:2118-2125(1989).  
RN [2]  
RP SEQUENCE OF 226-233; 256-267; 332-341; 667-677 AND 898-908.  
RX MEDLINE: 86287311.  
RA WEIS J.J., FEARON D.T., KLICKSTEIN L.B., MONG W.W., RICHARDS S.A.,  
RA DE BRYN KOS A., SMITH J.A., WEIS J.H.;  
RL PROC. NATL. ACAD. SCI. U.S.A. 83:5639-5643(1986).  
CC -1- FUNCTION: RECEPTOR FOR COMPLEMENT C3D AND FOR THE EPSTEIN-BARR  
CC VIRUS ON HUMAN B-CELLS. PARTICIPATES IN B LYMPHOCYTES ACTIVATION.  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -1- TISSUE SPECIFICITY: MATURE B LYMPHOCYTES AND FOLLICULAR  
CC DENDRITIC CELLS OF THE SPLEEN  
CC -1- SIMILARITY: TO MOUSE CR2. CD21 IS A MEMBER OF THE REGULATORS OF  
CC COMPLEMENT ACTIVATION (RCA) FAMILY.  
CC -1- SIMILARITY: CONTAINS 15 SUSHI (SCR) REPEATS.  
CC -1- DATABASE: NAME-PROV; NOTE-CD guide CD21 entry;  
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd21.htm".  
DR EMBL: M26004; G181940; -  
DR EMBL: M26016; G181938; -  
DR EMBL: M24007; G181938; JOINED.  
DR EMBL: M24008; G181938; JOINED.  
DR EMBL: M24009; G181938; JOINED.  
DR EMBL: M24010; G181938; JOINED.  
DR EMBL: M24011; G181938; JOINED.  
DR EMBL: M26008; G181938; JOINED.  
DR EMBL: M26010; G181938; JOINED.  
DR EMBL: M26011; G181938; JOINED.  
DR EMBL: M26012; G181938; JOINED.  
DR EMBL: M26013; G181938; JOINED.  
DR EMBL: M26014; G181938; JOINED.  
DR EMBL: M26015; G181938; JOINED.  
DR PIR: A32036; A32036.  
DR PIR: A24319; A24319.  
DR PIR: B24319; B24319.  
DR PIR: C24319; C24319.  
DR PIR: D24319; D24319.  
DR PIR: E24319; E24319.  
DR PIR: F24319; F24319.  
DR HSSP: P08603; 1HF1.  
DR MIM: 120650; -  
KM COMPLEMENT PATHWAY; GLYCOPROTEIN; TRANSMEMBRANE; REPEAT; SIGNAL;  
KW RECEPTOR; SUSHI.  
FT SIGNAL 1 20  
FT CHAIN 21 1033  
FT DOMAIN 21 971  
FT TRANSMEM 972 999  
FT DOMAIN 1000 1033  
FT DOMAIN 22 971  
FT DOMAIN 22 83  
FT REPEAT 90 147  
FT REPEAT 153 211  
FT REPEAT 214 272  
FT REPEAT 275 343  
FT REPEAT 350 407  
FT REPEAT 409 467  
FT REPEAT 470 523  
FT REPEAT 526 594  
FT REPEAT 601 658  
FT REPEAT 661 715  
FT REPEAT 718 780  
FT REPEAT 787 844  
FT REPEAT 850 908  
FT REPEAT 911 969  
FT DISULFID 23 65  
FT DISULFID 51 82  
FT DISULFID 91 132  
FT DISULFID 118 146  
FT DISULFID 154 197  
FT DISULFID 183 210

FT DISULFID 215 256 BY SIMILARITY.  
FT DISULFID 242 271 BY SIMILARITY.  
FT DISULFID 276 325 BY SIMILARITY.  
FT DISULFID 305 342 BY SIMILARITY.  
FT DISULFID 351 393 BY SIMILARITY.  
FT DISULFID 379 406 BY SIMILARITY.  
FT DISULFID 410 453 BY SIMILARITY.  
FT DISULFID 439 466 BY SIMILARITY.  
FT DISULFID 471 509 BY SIMILARITY.  
FT DISULFID 495 522 BY SIMILARITY.  
FT DISULFID 527 576 BY SIMILARITY.  
FT DISULFID 556 593 BY SIMILARITY.  
FT DISULFID 602 644 BY SIMILARITY.  
FT DISULFID 630 657 BY SIMILARITY.  
FT DISULFID 662 699 BY SIMILARITY.  
FT DISULFID 685 714 BY SIMILARITY.  
FT DISULFID 719 762 BY SIMILARITY.  
FT DISULFID 748 779 BY SIMILARITY.  
FT DISULFID 788 830 BY SIMILARITY.  
FT DISULFID 816 843 BY SIMILARITY.  
FT DISULFID 851 894 BY SIMILARITY.  
FT DISULFID 912 907 BY SIMILARITY.  
FT DISULFID 941 955 BY SIMILARITY.  
FT DISULFID 121 121 BY SIMILARITY.  
FT CARBOHYD 127 127 POTENTIAL.  
FT CARBOHYD 294 294 POTENTIAL.  
FT CARBOHYD 372 372 POTENTIAL.  
FT CARBOHYD 492 492 POTENTIAL.  
FT CARBOHYD 623 623 POTENTIAL.  
FT CARBOHYD 682 682 POTENTIAL.  
FT CARBOHYD 800 800 POTENTIAL.  
FT CARBOHYD 823 823 POTENTIAL.  
FT CARBOHYD 861 861 POTENTIAL.  
FT CARBOHYD 911 911 POTENTIAL.  
FT CONFLICT 667 667 Q -> D (IN REF. 2).  
FT CONFLICT 902 902 Q -> G (IN REF. 2).  
FT CONFLICT 906 906 H -> L (IN REF. 2).  
SO SEQUENCE 1033 AA; 112973 MW; 7032F558 CRC32;  
  
Query Match 45.7%; Score 193; DB 1; Length 1033;  
Best local Similarity 52.2%; Pred. No. 1.7e-14;  
Matches 36; Conservative 7; Mismatches 26; Indels 0; Gaps 0;  
  
QY 2 IPCGLPPTINGDFISTNNENHYGVSVTYRCNPGSGGRKVFELVGEPSITYCTSNDDVG 61  
DB 525 ITCPPPPVYINGAHGTSSLEDFFYGTVTYTCNPGRGVGFELIGESTIRCTSNDDQERG 584  
QY 62 IWSGPAPQC 70  
DB 585 TWSGPAPLC 593  
  
RESULT 3  
CR2\_MOUSE STANDARD; PRT: 1025 AA.  
AC P19070;  
DT 01-NOV-1990 (REL. 16, CREATED)  
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)  
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
DE COMPLEMENT RECEPTOR TYPE 2 PRECURSOR (CR2) (COMPLEMENT C3D RECEPTOR).  
GN CR2.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA.  
OC EUTHERIA; RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BALB/C;  
RX MEDLINE: 90229735.  
RA FINGEROTH J.D.;  
RL J. IMMUNOL. 144:3458-3467(1990).  
RN [2]

RP SEQUENCE OF 12-1025 FROM N.A.  
 RA MEDLINE; 91010789.  
 RA MOLINA H., KINOSHITA T., INOUE K., CAREL J.C., HOLERS V.M.;  
 J. IMMUNOL. 145:2974-2983(1990).  
 RN [3]  
 RP SEQUENCE OF 343-401 AND 991-1025 FROM N.A.  
 RA MEDLINE; 89098890.  
 RA FINGEROTH J.D., BENEDICT M.A., LEVY D.N., STROMINGER J.L.;  
 PROC. NATL. ACAD. SCI. U.S.A. 86:242-246(1989).  
 RN [4]  
 RP SEQUENCE OF 289-1025 FROM N.A.  
 RA MEDLINE; 89381350.  
 RA KURTZ C.B., PAUL M.S., AEGERTER M., WEIS J.J., WEIS J.H.;  
 J. IMMUNOL. 143:2058-2067(1989).  
 RL -1- FUNCTION: RECEPTOR FOR COMPLEMENT C3D. PARTICIPATES IN B  
 LYMPHOCYTES ACTIVATION.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: B LYMPHOCYTES.  
 CC -1- SIMILARITY: TO HUMAN C3D/EBV RECEPTOR (CD21).  
 CC -1- SIMILARITY: CONTAINS 15 SUSHI (SCR) REPEATS.  
 DR EMBL; M81083; G192687; -;  
 DR EMBL; M35684; G192688; -;  
 DR EMBL; M61132; G192693; -;  
 DR EMBL; M35685; G192695; ALT.SEO.  
 DR EMBL; M29281; G387131; -;  
 DR PIR; A43526; A43526.  
 DR HSSP; P08603; IHCC.  
 DR MGD; MGI:88489; CR2.  
 KW COMPLEMENT PATHWAY; GLYCOPROTEIN; TRANSMEMBRANE; REPEAT; SIGNAL;  
 KM RECEPTOR; SUSHI.  
 FT SIGNAL 1 11  
 FT CHAIN 12 1025  
 FT DOMAIN 12 963  
 FT TRANSMEM 964 990  
 FT DOMAIN 991 1025  
 FT DOMAIN 13 959  
 FT REPEAT 13 74  
 FT REPEAT 81 139  
 FT REPEAT 145 203  
 FT REPEAT 206 264  
 FT REPEAT 267 335  
 FT REPEAT 342 399  
 FT REPEAT 401 458  
 FT REPEAT 462 515  
 FT REPEAT 518 586  
 FT REPEAT 593 649  
 FT REPEAT 653 705  
 FT REPEAT 708 769  
 FT REPEAT 777 834  
 FT REPEAT 840 898  
 FT REPEAT 901 959  
 FT DISULFID 14 56  
 FT DISULFID 42 73  
 FT DISULFID 82 124  
 FT DISULFID 110 138  
 FT DISULFID 146 189  
 FT DISULFID 175 202  
 FT DISULFID 207 248  
 FT DISULFID 234 263  
 FT DISULFID 268 317  
 FT DISULFID 297 334  
 FT DISULFID 343 385  
 FT DISULFID 371 398  
 FT DISULFID 402 445  
 FT DISULFID 431 458  
 FT DISULFID 463 501  
 FT DISULFID 487 514  
 FT DISULFID 519 568  
 FT DISULFID 548 585  
 FT DISULFID 594 636  
 FT DISULFID 622 649  
 FT DISULFID 654 689  
 FT DISULFID 675 704

FT DISULFID 709 752 BY SIMILARITY.  
 FT DISULFID 738 769 BY SIMILARITY.  
 FT DISULFID 778 820 BY SIMILARITY.  
 FT DISULFID 806 833 BY SIMILARITY.  
 FT DISULFID 841 884 BY SIMILARITY.  
 FT DISULFID 870 897 BY SIMILARITY.  
 FT DISULFID 902 945 BY SIMILARITY.  
 FT DISULFID 931 958 BY SIMILARITY.  
 FT CARBOHYD 77 77 POTENTIAL.  
 FT CARBOHYD 113 113 POTENTIAL.  
 FT CARBOHYD 276 276 POTENTIAL.  
 FT CARBOHYD 316 316 POTENTIAL.  
 FT CARBOHYD 364 364 POTENTIAL.  
 FT CARBOHYD 380 380 POTENTIAL.  
 FT CARBOHYD 484 484 POTENTIAL.  
 FT CARBOHYD 527 527 POTENTIAL.  
 FT CARBOHYD 615 615 POTENTIAL.  
 FT CARBOHYD 639 639 POTENTIAL.  
 FT CARBOHYD 694 694 POTENTIAL.  
 FT CARBOHYD 754 754 POTENTIAL.  
 FT CARBOHYD 790 790 POTENTIAL.  
 FT CARBOHYD 813 813 POTENTIAL.  
 FT CARBOHYD 823 823 POTENTIAL.  
 FT CARBOHYD 851 851 POTENTIAL.  
 FT CARBOHYD 901 901 POTENTIAL.  
 FT CONFLICT 289 291 YGS -> EFR (IN REF. 4).  
 FT CONFLICT 306 306 S -> T (IN REF. 2).  
 FT CONFLICT 520 520 P -> A (IN REF. 2).  
 FT CONFLICT 962 963 MISSING (IN REF. 4).  
 SQ SEQUENCE 1025 AA; 112994 MW; DA043B62 CRC32;

Query Match 41.7%; Score 176; DB 1; Length 1025;  
 Best Local Similarity 45.1%; Pred No. 1.5e-12;  
 Matches 32; Conservative 9; Mismatches 30; Indels 0; Gaps 0;  
 QY 2 IPCGLPTTNGDPISTNRENFHSGVYTYRCNPGSGRKVFELVGPSTYCTSDNDVQ 61  
 Db 266 ILCPPEPPVRRSHSTGSESVNPYGTVTYCDPSPEKGVSTLIGKTIINCTIGSGKGT 325  
 QY 62 IWSGPAPCYVL 72  
 Db 326 IWSGPAPCYVL 336  
 RESULT 4  
 DAF\_PONPY STANDARD; PRT; 340 AA.  
 ID DAF\_PONPY  
 AC P49457;  
 DT 01-FEB-1996 (REL. 33, CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE COMPLEMENT DECAT-ACCELERATING FACTOR (CD55) (FRAGMENT).  
 GN DAF OR CD55.  
 OS PONGO PYGMAEUS (ORANGUTAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 94110622.  
 RA NICKELLS M.W., ALVAREZ J.I., LOBLIN D.M., ATRINSON J.P.;  
 J. IMMUNOL. 152:676-685(1994).  
 RL -1- FUNCTION: THIS PROTEIN RECOGNIZES C4B AND C3B FRAGMENTS THAT  
 CONDENSE WITH CELL-SURFACE HYDROXYL OR AMINO GROUPS WHEN NASCENT  
 C4B AND C3B ARE LOCALLY GENERATED DURING C4 AND C3 POLYPEPTIDES  
 INTERFERES WITH THEIR ABILITY TO CATALYZE THE CONVERSION OF C2 AND  
 FACTOR B TO ENZYMATICALLY ACTIVE C2A AND BB AND THEREBY PREVENTS  
 THE FORMATION OF C4B2A AND C3BB. THE AMPLIFICATION CONVERTASES OF  
 THE COMPLEMENT CASCADE (BY SIMILARITY).  
 CC -1- SUBUNIT: MONOMER (MAJOR FORM) AND NON-DISULFIDE-LINKED, COVALENT  
 HOMODIMER (MINOR FORM).  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.

CC -1- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR  
CC FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE  
CC ACTIVE SITE ON SCR3 (BY SIMILARITY).  
CC -1- P.T.M. THE SER/THR-RICH DOMAIN IS HEAVILY O-GLYCOSYLATED.  
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF DAF (DAF-2, SHOWN HERE, AND  
CC DAF-1) ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.  
CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.  
CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION  
CC (RCA) FAMILY.  
DR EMBL: S67775; G459390; -  
KM COMPLEMENT PATHWAY: PLASMA; GLYCOPROTEIN; MEMBRANE; REPEAT;  
KM ALTERNATIVE SPLICING; GPI-ANCHOR; SUSHI.  
FT NON-TER 1  
FT CHAIN 1 312  
FT PROPEP 313 340 COMPLEMENT DECAV-ACCELERATING FACTOR.  
FT DOMAIN <1 243 REMOVED IN MATURE FORM (BY SIMILARITY).  
FT REPEAT <1 54 4 X SUSHI (SCR) REPEATS.  
FT REPEAT 56 118 SUSHI 1.  
FT REPEAT 121 180 SUSHI 2.  
FT REPEAT 183 243 SUSHI 3.  
FT DOMAIN 246 315 SUSHI 4.  
FT DISULFID 24 53 SER/THR-RICH.  
FT DISULFID 57 104 BY SIMILARITY.  
FT DISULFID 86 117 BY SIMILARITY.  
FT DISULFID 122 163 BY SIMILARITY.  
FT DISULFID 149 179 BY SIMILARITY.  
FT DISULFID 184 226 BY SIMILARITY.  
FT DISULFID 212 242 BY SIMILARITY.  
FT CARBOHYD 54 54 POTENTIAL.  
FT CARBOHYD 107 107 POTENTIAL.  
FT LIPID 312 312 GPI-ANCHOR (BY SIMILARITY).  
SQ SEQUENCE 340 AA; 37180 MW; 5CAC18F8 CRC32;

Query Match 37.9%; Score 160; DB 1; Length 340;  
Best Local Similarity 46.4%; Pred. No. 3e-11;  
Matches 32; Conservative 9; Mismatches 20; Indels 8; Gaps 3;

QY 2 IPCGLPTINGDFISNRENFHGVTVTCNPGSGRKVELGHPSTYCTSDNDQVG 61  
DB 182 IYCAPPOINDG-ITQKRHDHGRGROSITYACNKG-----YIMGHSHITCYNDDE-G 233  
QY 62 IWSGPAPQC 70  
DB 234 EWSGPPECC 242

RESULT 5  
DAF\_HUMAN STANDARD; PRT; 381 AA.  
ID DAF\_HUMAN  
AC P08174; P09679;  
DT 01-AUG-1988 (REL. 08, CREATED)  
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE COMPLEMENT DECAV-ACCELERATING FACTOR PRECURSOR (CD55).  
GN DAF OR CD55.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN 1  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 87115845.  
RA CARAS I.W., DAVITZ M.A., RHEE L., WEDDELL G., MARTIN D.W. JR.,  
RA NUSSENZWEIG V.;  
RL NATURE 325:545-549(1987).  
RN 2  
RP SEQUENCE OF 6-381 FROM N.A. (DAF-2 FORM).  
RX MEDLINE: 87175602.  
RA MEDO M.E., LUBLIN D.M., HOLERS V.M., AYERS D.J., GETTY R.R.,  
RA LEXKAM J.F., ATKINSON J.P., TYKOCINSKI M.L.;  
RL PROC. NATL. ACAD. SCI. U.S.A. 84:2007-2011(1987).  
RN 3  
RP SEQUENCE OF 1-100 FROM N.A.

RX MEDLINE: 91271256.  
RA EMBLON U.K., RAVI L., MEDO M.E.;  
RL PROC. NATL. ACAD. SCI. U.S.A. 88:4675-4679(1991).  
RN [4]  
RP SEQUENCE OF 35-46.  
RC TISSUE-ORINE;  
RX MEDLINE: 91291869.  
RA NAKANO Y., SUGITA Y., ISHIKAWA Y., CHOI N.-H., TOBE T., TOMITA M.;  
RL BIOCHIM. BIOPHYS. ACTA 1074:326-330(1991).  
RN [5]  
RP GPI-ANCHOR.  
RX MEDLINE: 91093238.  
RA MORAN P., RAAB H., KOHR W.J., CARAS I.W.;  
RL J. BIOL. CHEM. 266:1250-1257(1991).  
RN [6]  
RP DISULFIDE BONDS IN SUSHI DOMAINS.  
RX MEDLINE: 92303034.  
RA NAKANO Y., SUMIDA K., KIKUTA N., MIURA N.-H., TOBE T., TOMITA M.;  
RL BIOCHIM. BIOPHYS. ACTA 1116:235-240(1992).  
RN [7]  
RP FUNCTION AS A ECHOVIRUS RECEPTOR.  
RX MEDLINE: 95045399.  
RA WARD T., PIPKIN P.A., CLARKSON N.A., STONE D.M., MINOR P.D.,  
RA ALMOND J.W.;  
RN EMBL J. 13:5070-5074(1994).  
RN [8]  
RP VARIANT BLOOD GROUP DR(A-).  
RX MEDLINE: 94325573.  
RA LUBLIN D.M., MALLINSON G., POOLE J., REID M.E., THOMPSON E.S.,  
RA FERDMAN B.R., TELEN M.J., ANSTEE D.J., TANNER M.J.A.;  
RL BLOOD 84:1276-1282(1994).  
RN [9]  
RP FUNCTION: THIS PROTEIN RECOGNIZES C4B AND C3B FRAGMENTS THAT  
CC CONDENSE WITH CELL-SURFACE HYDROXYL OR AMINO GROUPS WHEN NASCENT  
CC C4B AND C3B ARE LOCALLY GENERATED DURING C4 AND C3 ACTIVATION.  
CC INTERACTION OF DAF WITH CELL-ASSOCIATED C4B AND C3B POLYPEPTIDES  
CC INTERFERES WITH THEIR ABILITY TO CATALYZE THE CONVERSION OF C2 AND  
CC FACTOR B TO ENZYMATICALLY ACTIVE C2A AND BB AND THEREBY PREVENTS  
CC THE FORMATION OF C4B2A AND C3BBB, THE AMPLIFICATION CONVERTASES OF  
CC THE COMPLEMENT CASCADE.  
CC -1- FUNCTION: ALSO ACTS AS THE RECEPTOR FOR ECHOVIRUS 7 AND RELATED  
CC VIRUSES (ECHOVIRUSES 13, 21, 29 AND 33).  
CC -1- SUBUNIT: MONOMER (MAJOR FORM) AND NON-DISULFIDE-LINKED, COVALENT  
CC HOMODIMER (MINOR FORM).  
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.  
CC -1- TISSUE SPECIFICITY: EXPRESSED ON THE PLASMA MEMBRANES OF ALL CELL  
CC TYPES THAT ARE IN INTIMATE CONTACT WITH PLASMA COMPLEMENT  
CC PROTEINS. IT IS ALSO FOUND ON THE SURFACES OF EPITHELIAL CELLS  
CC LINING EXTRACELLULAR COMPARTMENTS, AND VARIANTS OF THE MOLECULE  
CC ARE PRESENT IN BODY FLUIDS AND IN EXTRACELLULAR MATRIX.  
CC -1- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR  
CC FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE  
CC ACTIVE SITE ON SCR3 (BY SIMILARITY).  
CC -1- P.T.M. THE SER/THR-RICH DOMAIN IS HEAVILY O-GLYCOSYLATED.  
CC -1- POLYMORPHISM: DAF IS RESPONSIBLE FOR THE CROMER BLOOD GROUP  
CC SYSTEM. IT CONSISTS OF AT LEAST SEVEN HIGH-INCIDENCE (CRA),  
CC TC(A), DR(A), ES(A), WE(SB), UMC, AND IFC) AND LOW-INCIDENCE  
CC (TC(B), TC(C), AND WE(SA)) ANTIGENS THAT RESIDE ON DAF. IN THE  
CC CROMER PHENOTYPES DR(A-) AND INAB THERE IS REDUCED OR ABSENT  
CC EXPRESSION OF DAF, RESPECTIVELY. IN THE CASE OF THE DR(A-) PHENOTYPE,  
CC A SINGLE NUCLEOTIDE SUBSTITUTION WITHIN EXON 5 ACCOUNTS  
CC FOR TWO CHANGES: A SIMPLE AMINO ACID SUBSTITUTION THAT IS THE  
CC BASIS OF THE ANTIGENIC VARIATION, AND AN ALTERNATIVE SPLICING  
CC EVENT THAT UNDERLIES THE DECREASED EXPRESSION OF DAF IN THIS  
CC PHENOTYPE.  
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF DAF (DAF-2, SHOWN HERE, AND  
CC DAF-1) ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.  
CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.  
CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION  
CC (RCA) FAMILY.  
CC -1- DATABASE: NAME-PROW: NOTE-CD guide CD55 entry;  
CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd55.htm".  
DR EMBL: M31516; G181468; -  
DR EMBL: M30142; G181465; -

DR EMBL; M15799; G181463; -  
 DR EMBL; M64653; G181476; -  
 DR EMBL; M64356; G181476; JOINED.  
 DR EMBL; S72858; G639600; -  
 DR PIR; B26359; B26359.  
 DR PIR; A26359; A26359.  
 DR PIR; S16187; S16187.  
 DR PIR; A39101; A39101.  
 DR PIR; S23138; S23138.  
 DR HSP; P08603; IHEI.  
 DR MM; 125240; -  
 KM COMPLEMENT PATHWAY; PLASMA; GLYCOPROTEIN; MEMBRANE; REPEAT;  
 KM ALTERNATIVE SPLICING; GPI-ANCHOR; SIGNAL; SUSHI; POLYMORPHISM;  
 KM BLOOD GROUP ANTIGEN.  
 FT SIGNAL 1 34  
 FT CHAIN 35 353  
 FT PROPEP 354 381  
 FT DOMAIN 35 284  
 FT REPEAT 35 95  
 FT REPEAT 97 159  
 FT REPEAT 162 221  
 FT REPEAT 224 284  
 FT DOMAIN 287 356  
 FT DISULFID 36 81  
 FT DISULFID 65 94  
 FT DISULFID 98 145  
 FT DISULFID 129 158  
 FT DISULFID 163 204  
 FT DISULFID 190 220  
 FT DISULFID 225 267  
 FT DISULFID 253 283  
 FT CARBOHYD 95 95  
 FT LIPID 353 353  
 FT VARIANT 52 52  
 FT VARIANT 52 52  
 FT VARIANT 82 82  
 FT VARIANT 199 199  
 FT VARIANT 227 227  
 FT VARSPLIC 362 381  
 FT CONFLICT 80 80  
 FT CONFLICT 85 85  
 FT SEQUENCE 381 AA; 41388 MM; 114271A8 CRC32;  
 Query Match 37.7%; Score 159; DB 1; Length 381;  
 Best Local Similarity 47.8%; Pred. No. 4.5e-11;  
 Matches 33; Conservative 8; Mismatches 20; Indels 8; Gaps 3;  
 QY 2 IPCGLPPTITNDGFISTRENFHGSVVTYRCNPGSGRKVFELVGPSTYCTSNDDQVG 61  
 DB 223 IYCAPAPQIDNG-IIOGSRDHGYRGSTVYACNKG-----FTMIGHSITCTVANNDE-G 274  
 QY 62 IWSGPAPOC 70  
 DB 275 EMSGPPPEC 283  
 RESULT 5  
 ID DAF1\_MOUSE STANDARD; PRT; 390 AA.  
 AC Q61475; Q61397;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE COMPLEMENT DECAT-ACCELERATING FACTOR, GPI-ANCHORED PRECURSOR  
 DE (DAF-GPI).  
 GN DAF1.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUHERIA; RODENTIA.  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN-C57BL/6J; TISSUE-TESTIS;  
 RX MEDLINE; 95403982.  
 RA SPICER A.P., SELDIN M.F., GENDLER S.J.;  
 RL J. IMMUNOL. 155:3079-3091(1995).  
 RN [2]  
 RP SEQUENCE OF 7-390 FROM N.A.  
 RC STRAIN-BALB/C; TISSUE-SPLEEN;  
 RX MEDLINE; 96362213.  
 RA FUKUOKA Y., YASUI A., OKADA N., OKADA H.;  
 RL INT. IMMUNOL. 8:379-385(1996).  
 CC -1- FUNCTION: PROTECTION OF CELLS FROM COMPLEMENT-MEDIATED DAMAGE (BY  
 CC SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.  
 CC -1- TISSUE SPECIFICITY: BRAIN, SECRETORY EPITHELIA, SKELETAL MUSCLE,  
 CC LIVER, TESTES, THYMUS, SPLEEN AND LYMPH NODE.  
 CC -1- DOMAIN: THE FIRST SUSHI DOMAIN (SC1) IS NOT NECESSARY FOR THE  
 CC ACTIVE SITE ON SC3 (BY SIMILARITY).  
 CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.  
 CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION  
 CC (RCA) FAMILY.  
 DR EMBL; L41366; G886335; -  
 DR EMBL; D63679; D1010476; -  
 DR MGI; MGI:104850; DAF1.  
 KM COMPLEMENT PATHWAY; GLYCOPROTEIN; MEMBRANE; REPEAT; GPI-ANCHOR;  
 KM SIGNAL; SUSHI.  
 FT SIGNAL 1 34  
 FT CHAIN 35 361  
 FT PROPEP 362 390  
 FT DOMAIN 35 285  
 FT REPEAT 35 95  
 FT REPEAT 97 159  
 FT REPEAT 162 221  
 FT REPEAT 224 283  
 FT DOMAIN 288 364  
 FT DISULFID 65 94  
 FT DISULFID 98 145  
 FT DISULFID 129 158  
 FT DISULFID 163 204  
 FT DISULFID 190 220  
 FT DISULFID 225 267  
 FT DISULFID 253 284  
 FT CARBOHYD 187 187  
 FT CARBOHYD 262 262  
 FT LIPID 361 361  
 FT CONFLICT 7 7  
 FT CONFLICT 9 9  
 FT CONFLICT 83 83  
 FT CONFLICT 91 91  
 FT CONFLICT 135 135  
 FT CONFLICT 173 173  
 FT CONFLICT 180 180  
 FT SEQUENCE 390 AA; 42618 MM; ADBDB34 CRC32;  
 Query Match 35.2%; Score 148.5; DB 1; Length 390;  
 Best Local Similarity 42.3%; Pred. No. 7.5e-10;  
 Matches 33; Conservative 8; Mismatches 26; Indels 11; Gaps 3;  
 QY 2 IPCGLPPTITNDGFISTRENFHGSVVTYRCNPGSGRKVFELVGPSTYCTSNDDQVG 61  
 DB 223 IHCEPPKINNG-IMRGESDSYTSQVVTYSCDNG-----FLVGNASTYCTVSKSDVG 275  
 QY 62 IWSGPAPOC 75  
 DB 276 QWSSPPPECIEKSKVPTK 293  
 RESULT 7  
 ID CABP\_RAT STANDARD; PRT; 558 AA.  
 AC Q61475; Q61397;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE COMPLEMENT DECAT-ACCELERATING FACTOR, GPI-ANCHORED PRECURSOR  
 DE (DAF-GPI).  
 GN DAF1.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUHERIA; RODENTIA.  
 RN [1]





```

DR EMBL: Y00651: G34505: -
DR EMBL: S51940: G262938: -
DR EMBL: M58050: G180137: -
DR EMBL: A18885: G512457: -
DR EMBL: S63879: E91388: -
DR PIR: S01896: S01896.
DR MIM: 120920: -
KW COMPLEMENT PATHWAY; GLYCOPROTEIN; TRANSMEMBRANE; REPEAT; SIGNAL;
KW SUSHI: ALTERNATIVE SPLICING.
FT SIGNAL 1 34
FT CHAIN 35 377
FT DOMAIN 35 328
FT TRANSMEM 329 351
FT DOMAIN 352 377
FT DOMAIN 35 284
FT REPEAT 35 95
FT REPEAT 98 158
FT REPEAT 161 224
FT REPEAT 227 284
FT DOMAIN 287 311
FT DISULFID 35 80
FT DISULFID 64 94
FT DISULFID 99 141
FT DISULFID 127 157
FT DISULFID 162 210
FT DISULFID 191 223
FT DISULFID 228 270
FT DISULFID 256 283
FT CARBOHYD 83 83
FT CARBOHYD 114 114
FT CARBOHYD 273 273
FT VASAPPLIC 286 300
FT VASAPPLIC 340 352
FT VASAPPLIC 353 361
FT FT 362 377
FT FT 362 377
FT FT 362 377
FT FT 340 361
FT FT 339 339
SO SEQUENCE 377 AA; 42247 MW; 7EE9E38E CRC32;

Query Match 31.4%; Score 132.5; DB 1; Length 377;
Best Local Similarity 40.0%; Pred. No. 5e-08; Matches 28; Indels 5; Gaps 2;
Matches 28; Conservative 9;

QY 1 RIPCLPTITNGDITINRNFHNGSVYTYCNGSGRKFVLEGPSITCSDPDQV 60
DB 159 KVLCTPPKINKGKHTFSEVEYFELIDAVTYSQDP-APQDPDFSLIGESTITCGDN--- 214
QY 61 GIWSPGAPQC 70
DB 214 SVWSRADEC 223

RESULT 9
DAF2_MOUSE STANDARD: PRT: 407 AA.
AC Q61476;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE COMPLEMENT DECAT-ACCELERATING FACTOR, TRANSMEMBRANE PRECURSOR
DE (DAF-TM).
DAF2.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERAIA; ROSENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;

```

```

RX MEDLINE 95403982.
RA SPICER A.P., SELDIN M.F., GENDLER S.J.
RL J. IMMUNOL. 155:3079-3091(1995).
CC -1- FUNCTION: PROTECTION OF CELLS FROM COMPLEMENT-MEDIATED DAMAGE (BY SMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).
CC -1- TISSUE SPECIFICITY: TESTES, SPLEEN AND LYMPH NODE.
CC -1- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE ACTIVE SITE ON SCR3 (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.
CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION (RCA) FAMILY.
DR EMBL: L41365; G886338; -.
DR MGD: MGI:104849; DAF2.
KW COMPLEMENT PATHWAY; GLYCOPROTEIN; REPEAT; SIGNAL; SUSHI; TRANSMEMBRANE.
KW TRANSMEMBRANE.
FT FT SIGNAL 1 39 POTENTIAL.
FT FT CHAIN 40 407 COMPLEMENT DECA-ACCELERATING FACTOR, TRANSMEMBRANE.
FT FT TRANSSEM 369 389 EXTRACELLULAR (POTENTIAL).
FT FT DOMAIN 390 407 POTENTIAL.
FT FT DOMAIN 40 290 POTENTIAL.
FT FT REPEAT 40 100 4 X SUSHI (SCR) REPEATS.
FT FT REPEAT 102 164 SUSHI 1.
FT FT REPEAT 167 226 SUSHI 2.
FT FT REPEAT 229 290 SUSHI 3.
FT FT REPEAT 229 290 SUSHI 4.
FT FT DOMAIN 291 363 SER/THR-RICH (BY SIMILARITY).
FT FT DISULFID 70 99 BY SIMILARITY.
FT FT DISULFID 103 150 BY SIMILARITY.
FT FT DISULFID 134 163 BY SIMILARITY.
FT FT DISULFID 168 209 BY SIMILARITY.
FT FT DISULFID 195 225 BY SIMILARITY.
FT FT DISULFID 230 272 BY SIMILARITY.
FT FT DISULFID 258 289 BY SIMILARITY.
FT FT CARBOHYD 192 192 POTENTIAL.
FT FT CARBOHYD 267 267 POTENTIAL.
SQ SEQUENCE 407 AA; 44469 MW; BABF07E9 CRC32;

Query Match 30.9%; Score 130.5; DB 1; Length 407;
Best Local Similarity 38.9%; Pred. No. 9.2e-08;
Matches 26; Conservative 9; Mismatches 28; Indels 7; Gaps 2;

QY 4 CGLPPTTNGDFISTNRENFHYSWVYTRCNPGSGGRKVFELVGEPSIYCTSDNDVQGIW 63
DB 230 CPDPPKINDG-IMGESDSYKTSQVVIYSCDKG-----FLTFGNSITLYCVIYSDVQGM 282
QY 64 SGPAPOCIIPNK 75
DB 283 SSPPQCIIEESK 294

RESULT 10
DAF-CAVPO
DAF-CAVPO STANDARD; PRT; 507 AA.
AC Q6Q401; Q6Q402; Q6Q403; Q6Q404; Q6Q405; Q6Q406; P97354; P97255;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE COMPLEMENT DECA-ACCELERATING FACTOR PRECURSOR.
DAF.
OS CAVIA PORCELLUS (GUINEA PIG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUHERETIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HARTLEY; TISSUE=SPLEEN;
RX MEDLINE; 95403978.
RA NONAKA M., MIWA T., OKADA N., NONAKA M., OKADA H.;
RL J. IMMUNOL. 155:3037-3048(1995).
LC -1- FUNCTION: PREVENTS THE FORMATION AND/OR ACCELERATES THE

```



OC EUKARYOTA; METAFOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUHERIA; RODENTIA.  
 RN [1]  
 RX SEQUENCE FROM N.A.  
 RA MEDLINE: 88024997.  
 RA KRISTENSEN T., OGATA R.T., CHUNG L.P., REID K.B.M., TACK B.F.;  
 RL BIOCHEMISTRY 26:4668-4674(1987).  
 CC -1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT  
 CC ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR  
 CC (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT  
 CC ALSO ACCELERATES THE DEGRADATION OF THE C4BC2A COMPLEX (C3  
 CC CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA  
 CC CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM AMYLOID P COMPONENT.  
 CC -1- SUBUNIT: HOMOPOLYMER; NOT COVALENTLY LINKED. MOUSE LACKS THE  
 CC BETA CHAIN OF C4BP.  
 CC -1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.  
 CC -1- SIMILARITY: CONTAINS 6 SUSHI (SCR) REPEATS.  
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-44 IS THE INITIATOR.  
 DR EMBL: M17122; G309119; ALT\_INIT.  
 DR PIR: A27117; NEMSC4.  
 DR HSSP: P08603; 1HFI.  
 DR MGI: 882229; C4BP.  
 KW COMPLEMENT PATHWAY; PLASMA; GLYCOPROTEIN; REPEAT; SUSHI; SIGNAL.  
 FT SIGNAL 1 56  
 FT CHAIN 57 469 C4B-BINDING PROTEIN.  
 FT DOMAIN 57 414 6 X SUSHI (SCR) REPEATS.  
 FT REPEAT 57 116 SUSHI 1.  
 FT REPEAT 119 177 SUSHI 2.  
 FT REPEAT 180 241 SUSHI 3.  
 FT REPEAT 244 300 SUSHI 4.  
 FT REPEAT 302 356 SUSHI 5.  
 FT REPEAT 358 414 SUSHI 6.  
 FT DISULFID 58 103 BY SIMILARITY.  
 FT DISULFID 88 115 BY SIMILARITY.  
 FT DISULFID 120 160 BY SIMILARITY.  
 FT DISULFID 146 176 BY SIMILARITY.  
 FT DISULFID 181 223 BY SIMILARITY.  
 FT DISULFID 209 240 BY SIMILARITY.  
 FT DISULFID 245 287 BY SIMILARITY.  
 FT DISULFID 273 299 BY SIMILARITY.  
 FT DISULFID 303 343 BY SIMILARITY.  
 FT DISULFID 329 355 BY SIMILARITY.  
 FT DISULFID 359 400 BY SIMILARITY.  
 FT DISULFID 386 413 BY SIMILARITY.  
 FT CARBOHYD 74 74 POTENTIAL.  
 FT CARBOHYD 227 227 POTENTIAL.  
 FT CARBOHYD 275 275 POTENTIAL.  
 FT CARBOHYD 292 292 POTENTIAL.  
 FT CARBOHYD 366 366 POTENTIAL.  
 FT CARBOHYD 381 381 POTENTIAL.  
 FT CARBOHYD 428 428 POTENTIAL.  
 SO SEQUENCE 469 AA; 51551 MM; E05527AC CRC32;  
 Query Match 30.2%; Score 127.5; DB 1; Length 469;  
 Best Local Similarity 40.0%; Pred. No. 2.4e-07;  
 Matches 28; Conservative 9; Mismatches 24; Indels 9; Gaps 3;  
 Oy 2 IPCGPIITINGDISTRENH-YGSVYVTRCNGSGRGVFEVLYGSPSYICSNDOY 60  
 Db 179 VKCGPPPIISMKHSQT--EDYPPYHNGISYTCDBG-----FLVGSFPIGCTVYVNTYV 230  
 Oy 61 GWSGPAPQC 70  
 Db 231 PWSSSPPTC 240  
 RESULT 12  
 ID C4BP\_HUMAN STANDARD: PRT: 597 AA.  
 AC P04003: 23-OCT-1986 (REL. 02, CREATED)  
 DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE C4B-BINDING PROTEIN ALPHA CHAIN PRECURSOR (PROLINE-RICH PROTEIN)  
 DE (PRP).  
 GN C4BP OR C4BP.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAFOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE: LIVER.  
 RX MEDLINE: 90073699.  
 RA MATSUGUCHI T., OKAMURA S., ASO T., SATA T., NIHO Y.;  
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 165:138-144(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 91113199.  
 RA ASO T., OKAMURA S., MATSUGUCHI T., SAKAMOTO N., SATA T., NIHO Y.;  
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 174:222-227(1991).  
 RN [3]  
 RP SEQUENCE OF 9-81 FROM N.A.  
 RX MEDLINE: 88242821.  
 RA LINTIN S.J., LEWIN A.R., REID K.B.M.;  
 RL FEBS LETT. 232:328-332(1988).  
 RN [4]  
 RP SEQUENCE OF 203-288 FROM N.A.  
 RX MEDLINE: 86301119.  
 RA LINTIN S.J., REID K.B.M.;  
 RL FEBS LETT. 204:77-81(1986).  
 RN [5]  
 RP SEQUENCE OF 80-597 FROM N.A.  
 RX MEDLINE: 86025405.  
 RA CHUNG L.P., BENTLEY D.R., REID K.B.M.;  
 RL BIOCHEM. J. 230:133-141(1985).  
 RN [6]  
 RP SEQUENCE OF 49-88.  
 RX MEDLINE: 85296001.  
 RA CHUNG L.P., GAGNON J., REID K.B.M.;  
 RL MOL. IMMUNOL. 22:427-435(1985).  
 RN [7]  
 RP ELECTRON MICROSCOPY, 3-DIMENSIONAL STRUCTURE, AND LIGAND BINDING.  
 RX MEDLINE: 83221615.  
 RA DAHLBACK B., SMITH C.A., MUELLER-EBERHARD H.T.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 80:3461-3465(1983).  
 CC -1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT  
 CC ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR  
 CC (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT  
 CC ALSO ACCELERATES THE DEGRADATION OF THE C4BC2A COMPLEX (C3  
 CC CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA  
 CC CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM AMYLOID P COMPONENT.  
 CC -1- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS  
 CC OR 3 POSSIBLE SORCS: A 570 KD COMPLEX OF 7 ALPHA CHAINS AND 1 BETA  
 CC CHAIN, A 530-KD HOMOPOLYMER OF ALPHA CHAINS OR A 500-KD COMPLEX  
 CC OF 6 ALPHA CHAINS AND 1 BETA CHAIN. THE CENTRAL BODY OF THE  
 CC ALPHA CHAIN HOMOPOLYMER SUPPORTS TENTACLES, EACH WITH THE BINDING  
 CC SITE FOR C4B AT THE END.  
 CC -1- TISSUE SPECIFICITY: CHYLOMICRONS IN THE PLASMA.  
 CC -1- SIMILARITY: CONTAINS 8 SUSHI (SCR) REPEATS.  
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-17 IS THE INITIATOR.  
 DR EMBL: M31452; G190502; -.  
 DR EMBL: M62486; G190500; -.  
 DR EMBL: M62475; G190500; JOINED.  
 DR EMBL: M62476; G190500; JOINED.  
 DR EMBL: M62477; G190500; JOINED.  
 DR EMBL: M62478; G190500; JOINED.  
 DR EMBL: M62479; G190500; JOINED.  
 DR EMBL: M62480; G190500; JOINED.  
 DR EMBL: M62481; G190500; JOINED.  
 DR EMBL: M62482; G190500; JOINED.  
 DR EMBL: M62484; G190500; JOINED.  
 DR EMBL: M62485; G190500; JOINED.  
 DR EMBL: X07853; E3792; -

DR EMBL: X04284; E12637; -  
 DR EMBL: X04296; E12638; -  
 DR EMBL: X02865; G29565; -  
 DR PIR: A33568; NBHUC4.  
 DR HSSP: P08603; 1HCC.  
 DR MTM: 120830; -  
 KW COMPLEMENT PATHWAY; PLASMA; GLYCOPROTEIN; REPEAT; SUSHI; SIGNAL;  
 KM POLYMORPHISM.  
 FT SIGNAL 1 48  
 FT CHAIN 49 597  
 FT DOMAIN 49 539  
 FT REPEAT 49 109  
 FT REPEAT 112 171  
 FT REPEAT 174 235  
 FT REPEAT 238 295  
 FT REPEAT 298 361  
 FT REPEAT 364 423  
 FT REPEAT 425 481  
 FT REPEAT 483 539  
 FT DISULFID 50 96  
 FT DISULFID 81 108  
 FT DISULFID 113 154  
 FT DISULFID 140 170  
 FT DISULFID 175 217  
 FT DISULFID 203 234  
 FT DISULFID 239 281  
 FT DISULFID 267 294  
 FT DISULFID 299 348  
 FT DISULFID 332 360  
 FT DISULFID 365 409  
 FT DISULFID 426 468  
 FT DISULFID 454 480  
 FT DISULFID 484 525  
 FT DISULFID 511 538  
 FT DISULFID 546 546  
 FT DISULFID 558 558  
 FT CARBOHYD 221 221  
 FT CARBOHYD 506 506  
 FT CARBOHYD 528 528  
 FT VARIANT 92 92  
 FT VARIANT 357 357  
 SO SEQUENCE 597 AA; 67033 MW; DA7B4DC5 CRC32;  
 Q -> T.  
 Y -> H.  
 Query Match 29.5%; Score 124.5; DB 1; Length 597;  
 Best Local Similarity 38.6%; Pred. No. 7e-07;  
 Matches 27; Conservative 9; Mismatches 25; Indels 9; Gaps 3;  
 QY 2 IPCGLPTITNGDFISTNRENFH-YGSVVTYRCNPGSGGRKVFELVGEPSITCTSDNDQV 60  
 DB 173 VKCKPPPDINGR--HSGENFYAYGFSVITYSCDPR-----FSLJLHASISCTVENETI 224  
 QY 61 GINSAPAPOC 70  
 DB 225 GVMRPSPTC 234  
 RESULT 13  
 C4BP\_BOVIN STANDARD; PRT; 610 AA.  
 AC Q28065;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE C4B-BINDING PROTEIN ALPHA CHAIN PRECURSOR.  
 GN C4BPA.  
 OS BOS TAURUS (BOVINE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; ARTIODACTYLA.  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC TISSUE-LIVER;  
 RX MEDLINE; 95015909.  
 RA HILLARP A., THERN A., DAHLBAECH B.;  
 RL J. IMMUNOL. 153:4190-4199(1994).  
 CC -1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT  
 CC ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR  
 CC (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT  
 CC ALSO ACCELERATES THE DEGRADATION OF THE C4B2A COMPLEX (C3  
 CC CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA  
 CC CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM AMYLOID P COMPONENT.  
 CC -1- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS  
 CC (BY SIMILARITY).  
 CC -1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.  
 CC -1- SIMILARITY: CONTAINS 8 SUSHI (SCR) REPEATS.  
 DR EMBL: Z31693; G469118; -  
 KW COMPLEMENT PATHWAY; PLASMA; GLYCOPROTEIN; REPEAT; SUSHI; SIGNAL;  
 FT SIGNAL 1 48  
 FT CHAIN 49 610  
 FT DOMAIN 49 542  
 FT REPEAT 49 108  
 FT REPEAT 111 170  
 FT REPEAT 173 235  
 FT REPEAT 238 295  
 FT REPEAT 298 363  
 FT REPEAT 366 426  
 FT REPEAT 428 484  
 FT REPEAT 486 542  
 FT DISULFID 50 95  
 FT DISULFID 80 107  
 FT DISULFID 112 153  
 FT DISULFID 139 169  
 FT DISULFID 174 217  
 FT DISULFID 203 234  
 FT DISULFID 239 281  
 FT DISULFID 267 294  
 FT DISULFID 299 350  
 FT DISULFID 334 362  
 FT DISULFID 366 412  
 FT DISULFID 429 471  
 FT DISULFID 457 483  
 FT DISULFID 487 528  
 FT DISULFID 514 541  
 FT DISULFID 549 561  
 FT DISULFID 561 561  
 FT CARBOHYD 66 66  
 FT CARBOHYD 221 221  
 FT CARBOHYD 525 525  
 FT CARBOHYD 602 602  
 SO SEQUENCE 610 AA; 68886 MW; 39814F78 CRC32;  
 Query Match 29.4%; Score 124; DB 1; Length 610;  
 Best Local Similarity 35.8%; Pred. No. 8.1e-07;  
 Matches 24; Conservative 12; Mismatches 25; Indels 6; Gaps 1;  
 QY 4 CGLPPTITNGDFISTNRENFHGSVVTYRCNPGSGGRKVFELVGEPSITCTSDNDQV 63  
 DB 174 CEPPPTISNGRHNHGDDEFTYTGSSVYSCD-----RDFSMIGKASISCRVENKTIQW 227  
 QY 64 SGAPAPOC 70  
 DB 228 SPSPPSC 234  
 RESULT 14  
 F13B\_MOUSE STANDARD; PRT; 668 AA.  
 AC Q07968;  
 DT 01-FEB-1995 (REL. 31, CREATED)  
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE COAGULATION FACTOR XIII B CHAIN PRECURSOR (EC 2.3.2.13) (PROTEIN-  
DE GLUTAMINE GAMMA-GLUTAMYLTRANSFERASE B CHAIN) (TRANSGLUTAMINASE B  
CHAIN)  
GN F13B OR CF13B.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUKARYOTA; RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-B10.D2/OSN; TISSUE=LIVER;  
RX MEDLINE; 93224141.  
RA NAKAMURA M., MATSUDA Y., SHIROISHI T., MORIMAKI K., NONAKA M.,  
RL GENOMICS 15:535-542(1993).  
CC -1- FUNCTION: THE B CHAIN OF FACTOR XIII IS NOT CATALYTICALLY ACTIVE,  
BUT IS THOUGHT TO STABILIZE THE A SUBUNITS AND REGULATE THE RATE  
OF TRANSGLUTAMINASE FORMATION BY THROMBIN.  
CC -1- SUBUNIT: TETRAMER OF TWO A CHAINS AND TWO B CHAINS.  
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN LIVER AND KIDNEY.  
CC -1- SIMILARITY: CONTAINS 10 SUSHI (SCR) REPEATS.  
DR EMBL; D10071; G303652; -.  
DR PIR; A46013; A46013.  
DR MGI; 88379; CF13B.  
KW PLASMA; BLOOD COAGULATION; REPEAT; GLYCOPROTEIN; SIGNAL; SUSHI.  
FT SIGNAL 1 20  
FT CHAIN 21 668 BY SIMILARITY.  
FT DOMAIN 24 647 FACTOR XIII, B CHAIN.  
FT REPEAT 24 88 10 X SUSHI (SCR) REPEATS.  
FT REPEAT 90 147 SUSHI 1.  
FT REPEAT 152 209 SUSHI 2.  
FT REPEAT 212 268 SUSHI 3.  
FT REPEAT 273 328 SUSHI 4.  
FT REPEAT 335 390 SUSHI 5.  
FT REPEAT 395 451 SUSHI 6.  
FT REPEAT 453 516 SUSHI 7.  
FT REPEAT 523 579 SUSHI 8.  
FT REPEAT 581 647 SUSHI 9.  
FT REPEAT 647 766 SUSHI 10.  
FT DISULFID 25 76 BY SIMILARITY.  
FT DISULFID 59 87 BY SIMILARITY.  
FT DISULFID 91 135 BY SIMILARITY.  
FT DISULFID 118 146 BY SIMILARITY.  
FT DISULFID 153 197 BY SIMILARITY.  
FT DISULFID 180 208 BY SIMILARITY.  
FT DISULFID 213 255 BY SIMILARITY.  
FT DISULFID 241 267 BY SIMILARITY.  
FT DISULFID 274 316 BY SIMILARITY.  
FT DISULFID 302 327 BY SIMILARITY.  
FT DISULFID 336 378 BY SIMILARITY.  
FT DISULFID 364 389 BY SIMILARITY.  
FT DISULFID 396 439 BY SIMILARITY.  
FT DISULFID 425 450 BY SIMILARITY.  
FT DISULFID 454 505 BY SIMILARITY.  
FT DISULFID 486 515 BY SIMILARITY.  
FT DISULFID 524 567 BY SIMILARITY.  
FT DISULFID 553 578 BY SIMILARITY.  
FT DISULFID 582 636 BY SIMILARITY.  
FT DISULFID 616 646 BY SIMILARITY.  
FT CARBOHYD 162 162 POTENTIAL.  
FT CARBOHYD 545 545 POTENTIAL.  
SO SEQUENCE 668 AA; 76078 MW; 452345A0 CRC32;

Query Match 27.0%; Score 114; DB 1; Length 668;  
Best Local Similarity 34.2%; Pred. No. 1.3e-05;  
Matches 25; Conservative 9; Mismatches 21; Indels 18; Gaps 4;

OY 4 GCLPTINGDEISTNRENFHYGVSVTYRCNPGSGRAVFE---LVGEPSICTSNDQV 60  
DB 524 CASPVIINGDIVSAARFYENGSSVEYRC-----FDHFFQGSQNVYCV----- 570  
OY 61 GWSGPAFOCIIP 73  
DB 570 GWTTP-PSCLEP 581

RESULT 15  
ID CFAH\_MOUSE STANDARD; PRT: 1234 AA.  
AC P06909;  
DT 01-JAN-1988 (REL. 06, CREATED)  
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE COMPLEMENT FACTOR H PRECURSOR (PROTEIN BETA-1-H).  
GN CFAH.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUKARYOTA; RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 86233353.  
RA KRISTENSEN T., TACK B.F.;  
RL PROC. NATL. ACAD. SCI. U.S.A. 83:3963-3967(1986).  
RN [2]  
RP SEQUENCE OF 1-19 FROM N.A.  
RC STRAIN-BALB/C;  
RX MEDLINE; 90148935.  
RA MUNOZ-CANOVES P., TACK B.F., VIK D.P.;  
RL BIOCHEMISTRY 28:9891-9897(1989).  
RN [3]  
RP SEQUENCE OF 1-18 FROM N.A.  
RX MEDLINE; 90111033.  
RA NAKAMURA M., MATSUDA Y., SHIROISHI T., MORIMAKI K.,  
RL J. IMMUNOL. 144:358-362(1990).  
CC -1- FUNCTION: FACTOR H FUNCTIONS AS A COFACTOR IN THE INACTIVATION OF  
C3B BY FACTOR I AND ALSO INCREASES THE RATE OF DISSOCIATION OF THE  
C3BB COMPLEX (C3 CONVERTASE) AND THE (C3B)NB COMPLEX (C5  
CONVERTASE) IN THE ALTERNATIVE COMPLEMENT PATHWAY.  
CC -1- POLYMORPHISM: TWO CODOMINANT ALLELES OF FACTOR H ARE PRESENT IN  
MICE.  
CC EMBL; M12660; G387181; -.  
CC EMBL; J02891; G553926; -.  
CC EMBL; M31979; G193729; -.  
DR PIR; A26154; NEMSH.  
DR HSSP; P08603; IHFI.  
DR MGI; 88385; CFH.  
KW COMPLEMENT ALTERNATE PATHWAY; PLASMA; GLYCOPROTEIN; REPEAT; SUSHI.  
FT SIGNAL 1 18  
FT CHAIN 19 1234 BY SIMILARITY.  
FT DOMAIN 20 1224 COMPLEMENT FACTOR H.  
FT REPEAT 20 81 20 X SUSHI (SCR) REPEATS.  
FT REPEAT 84 142 SUSHI 1.  
FT REPEAT 145 206 SUSHI 2.  
FT REPEAT 209 263 SUSHI 3.  
FT REPEAT 266 321 SUSHI 4.  
FT REPEAT 324 386 SUSHI 5.  
FT REPEAT 388 443 SUSHI 6.  
FT REPEAT 447 506 SUSHI 7.  
FT REPEAT 508 565 SUSHI 8.  
FT REPEAT 568 623 SUSHI 9.  
FT REPEAT 628 684 SUSHI 10.  
FT REPEAT 689 744 SUSHI 11.  
FT REPEAT 751 803 SUSHI 12.  
FT REPEAT 807 862 SUSHI 13.  
FT REPEAT 866 932 SUSHI 14.  
FT REPEAT 935 990 SUSHI 15.  
FT REPEAT 993 1049 SUSHI 16.  
FT REPEAT 1052 1108 SUSHI 17.  
FT REPEAT 1113 1169 SUSHI 18.  
FT REPEAT 1171 1234 SUSHI 19.  
FT DISULFID 21 66 BY SIMILARITY.  
FT DISULFID 32 80 BY SIMILARITY.  
FT DISULFID 85 129 BY SIMILARITY.  
FT DISULFID 114 141 BY SIMILARITY.







GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 06:02:49 ; Search time 63.5 seconds  
(without alignments)  
65.160 Million cell updates/sec

Title: US-09-142-043-3

Perfect score: 422

Sequence: 1 RIPCGLPPTINGDFISTNR.....NDQVGIVSGPAPQCIPNK 75

Scoring table: BLOSUM62

Searched: 180763 segs, 55169189 residues

Database :

- 1: SPTRMBL\_8:\*
- 2: sp\_fungi:\*
- 3: sp\_human:\*
- 4: sp\_invertebrate:\*
- 5: sp\_mammal:\*
- 6: sp\_organelle:\*
- 7: sp\_phage:\*
- 8: sp\_plant:\*
- 9: sp\_bacteria:\*
- 10: sp\_rodent:\*
- 11: sp\_virus:\*
- 12: sp Vertebrate:\*
- 13: sp\_unclassified:\*
- 14: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	422	100.0	2039	2	016745	016745 homo sapien
2	422	100.0	2489	2	016744	016744 homo sapien
3	422	100.0	661	4	029331	029331 pan troglod
4	422	100.0	2014	4	029530	029530 pan troglod
5	373	88.4	1911	4	029528	029528 papio hamed
6	368	87.2	315	4	028770	028770 papio hamed
7	359	85.1	522	4	028769	028769 papio hamed
8	353	83.6	522	4	028797	028797 pan troglod
9	351	83.2	417	2	014079	014079 homo sapien
10	292	69.2	417	10	035520	035520 rattus norv
11	292	69.2	497	10	063612	063612 rattus norv
12	292	69.2	559	10	063135	063135 rattus norv
13	291	69.0	483	10	064735	064735 mus musculu
14	254	60.2	679	10	099254	099254 mus musculu
15	242	57.3	89	10	063129	063129 rattus norv
16	193	45.7	1032	2	013866	013866 homo sapien
17	193	45.7	1087	2	014212	014212 homo sapien
18	172	40.8	1045	4	046545	046545 ovis aries
19	159	37.7	347	2	P78361	P78361 homo sapien
20	148.5	35.2	390	10	P97732	P97732 mus musculu
21	146.5	34.7	222	4	019128	019128 pithecia pi
22	143.5	34.0	222	4	019122	019122 callimico g
23	141.5	33.5	336	4	062834	062834 saginus oe
24	141.5	33.5	314	4	062835	062835 saginus oe
25	141.5	33.5	378	4	062837	062837 saginus oe
26	141.5	33.5	377	4	062838	062838 saginus oe
27	141.5	33.5	222	4	019124	019124 saginus my
28	137.5	32.6	222	4	019123	019123 callithrix
29	137	32.5	533	10	008569	008569 cavia porce

30	135.5	32.1	222	4	019120	019120 aotus trivi
31	135.5	32.1	222	4	019125	019125 saimiri sci
32	134.5	31.9	363	4	002839	002839 sus scrofa
33	132.5	31.4	349	2	015429	015429 homo sapien
34	128.5	30.5	285	4	019126	019126 macaca fasc
35	128.5	30.5	285	4	019127	019127 macaca muli
36	128.5	30.5	369	4	P79138	P79138 cercopithe
37	127.5	30.2	285	4	019121	019121 papio hamed
38	121	28.7	579	10	060736	060736 mus musculu
39	114.5	27.1	379	10	P70105	P70105 cavia porce
40	114.5	27.1	339	10	P97261	P97261 cavia porce
41	114.5	27.1	337	10	P97263	P97263 cavia porce
42	114.5	27.1	236	10	P97264	P97264 cavia porce
43	114.5	27.1	267	10	P97939	P97939 cavia porce
44	114.5	27.1	318	10	P97262	P97262 cavia porce
45	114	27.0	365	10	086174	086174 mus musculu

## ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	2039 AA.
016745	016745			
AC	016745			
DT	01-NOV-1996 (TREMBLREL. 01, CREATED)			
DT	01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	COMPLEMENT RECEPTOR 1.			
OS	HOMO SAPIENS (HUMAN).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;			
OC	CATARRHINI; HOMINIDAE; HOMO.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 94065175.			
RA	VIK D.P., WONG W.W.;			
RT	*Structure of the gene for the F allele of complement receptor type 1 and sequence of the coding region unique to the S allele.*;			
RT	J. IMMUNOL. 151:6214-6224(1993).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	VIK D.P., WONG W.W.;			
RT	SUBMITTED (JUN-1993) TO EMBL/GENBANK/DBJ DATA BANKS.			
DR	EMBL; L17399; G306680; JOINED.			
DR	EMBL; L17409; G306680; JOINED.			
DR	EMBL; L17419; G306680; JOINED.			
DR	EMBL; L17420; G306680; JOINED.			
DR	EMBL; L17421; G306680; JOINED.			
DR	EMBL; L17422; G306680; JOINED.			
DR	EMBL; L17423; G306680; JOINED.			
DR	EMBL; L17381; G306680; JOINED.			
DR	EMBL; L17392; G306680; JOINED.			
DR	EMBL; L17393; G306680; JOINED.			
DR	EMBL; L17394; G306680; JOINED.			
DR	EMBL; L17395; G306680; JOINED.			
DR	EMBL; L17396; G306680; JOINED.			
DR	EMBL; L17387; G306680; JOINED.			
DR	EMBL; L17398; G306680; JOINED.			
DR	EMBL; L17418; G306680; JOINED.			
DR	EMBL; L17390; G306680; JOINED.			
DR	EMBL; L17400; G306680; JOINED.			
DR	EMBL; L17416; G306680; JOINED.			
DR	EMBL; L17417; G306680; JOINED.			
DR	EMBL; L17401; G306680; JOINED.			
DR	EMBL; L17402; G306680; JOINED.			
DR	EMBL; L17403; G306680; JOINED.			
DR	EMBL; L17404; G306680; JOINED.			
DR	EMBL; L17405; G306680; JOINED.			
DR	EMBL; L17406; G306680; JOINED.			
DR	EMBL; L17407; G306680; JOINED.			
DR	EMBL; L17408; G306680; JOINED.			
DR	EMBL; L17410; G306680; JOINED.			
DR	EMBL; L17411; G306680; JOINED.			

DR EMBL: L17412; G306680; JOINED.  
DR EMBL: L17413; G306680; JOINED.  
DR EMBL: L17414; G306680; JOINED.  
DR EMBL: L17415; G306680; JOINED.  
DR PRAM: PF00084; sush1; 30.  
SQ SEQUENCE 2039 AA; 223603 MW; A4972215 CRC32;

Query Match  
Best Local Similarity 100.0%; Score 422; DB 2; Length 2039;  
Pred. No. 2 6e-39;  
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RIPCGLPTTNGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSDNDQV 60  
|||||  
Db 163 RIPCGLPTTNGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSDNDQV 222  
|||||

QY 61 GWSGAPACIIPNK 75  
|||||  
Db 223 GWSGAPACIIPNK 237  
|||||

RESULT 2  
016744 PRELIMINARY; PRT; 2489 AA.  
AC 016744;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE COMPLEMENT RECEPTOR 1.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; HOMO.  
LN 11  
RN SEQUENCE FROM N.A.  
RP MEDLINE; 94065175.  
RA VIR D.P., WONG W.W.;  
RT "Structure of the gene for the F allele of complement receptor type 1  
and sequence of the coding region unique to the s allele.";  
RL J. IMMUNOL. 151:6214-6224(1993).  
RN 121  
RP SEQUENCE FROM N.A.  
RA VIR D.P., WONG W.W.;  
RL SUBMITTED (JUN-1993) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL: L17399; G451303; JOINED.  
DR EMBL: L17409; G451303; JOINED.  
DR EMBL: L17419; G451303; JOINED.  
DR EMBL: L17420; G451303; JOINED.  
DR EMBL: L17421; G451303; JOINED.  
DR EMBL: L17422; G451303; JOINED.  
DR EMBL: L17423; G451303; JOINED.  
DR EMBL: L17424; G451303; JOINED.  
DR EMBL: L17425; G451303; JOINED.  
DR EMBL: L17426; G451303; JOINED.  
DR EMBL: L17427; G451303; JOINED.  
DR EMBL: L17428; G451303; JOINED.  
DR EMBL: L17429; G451303; JOINED.  
DR EMBL: L17430; G451303; JOINED.  
DR EMBL: L17391; G451303; JOINED.  
DR EMBL: L17392; G451303; JOINED.  
DR EMBL: L17393; G451303; JOINED.  
DR EMBL: L17394; G451303; JOINED.  
DR EMBL: L17395; G451303; JOINED.  
DR EMBL: L17396; G451303; JOINED.  
DR EMBL: L17397; G451303; JOINED.  
DR EMBL: L17398; G451303; JOINED.  
DR EMBL: L17400; G451303; JOINED.  
DR EMBL: L17401; G451303; JOINED.  
DR EMBL: L17402; G451303; JOINED.  
DR EMBL: L17403; G451303; JOINED.

DR EMBL: L17405; G451303; JOINED.  
DR EMBL: L17406; G451303; JOINED.  
DR EMBL: L17407; G451303; JOINED.  
DR EMBL: L17408; G451303; JOINED.  
DR EMBL: L17410; G451303; JOINED.  
DR EMBL: L17411; G451303; JOINED.  
DR EMBL: L17412; G451303; JOINED.  
DR EMBL: L17413; G451303; JOINED.  
DR EMBL: L17414; G451303; JOINED.  
DR EMBL: L17415; G451303; JOINED.  
DR PRAM: PF00084; sush1; 37.  
SQ SEQUENCE 2489 AA; 272846 MW; 5869B6F9 CRC32;

Query Match  
Best Local Similarity 100.0%; Score 422; DB 2; Length 2489;  
Pred. No. 3 2e-39;  
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RIPCGLPTTNGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSDNDQV 60  
|||||  
Db 163 RIPCGLPTTNGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSDNDQV 222  
|||||

QY 61 GWSGAPACIIPNK 75  
|||||  
Db 223 GWSGAPACIIPNK 237  
|||||

RESULT 3  
029531 PRELIMINARY; PRT; 661 AA.  
AC 029531;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE COMPLEMENT RECEPTOR 1 PRECURSOR (FRAGMENT).  
GN CRI.  
OS PAN TROGLODYTES (CHIMPANZEE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; PAN.  
LN 11  
RN SEQUENCE FROM N.A.  
RP MEDLINE; 94292799.  
RA BIRMINGHAM D.J., SHEN X.P., HOURCADE D., NICKELIS M.W., ATKINSON J.P.;  
RT "Primary sequence of an alternatively spliced form of CRI. Candidate  
RT for the 75,000 M(r) complement receptor expressed on chimpanzee  
erythrocytes";  
RL J. IMMUNOL. 153:691-700(1994).  
DR EMBL: L24921; G557727; .  
DR PRAM: PF00084; sush1; 9.  
KW SIGNAL; ALTERNATIVE SPLICING.  
FT NON\_TER 1  
FT SIGNAL 1  
FT CHAIN <1 16 POTENTIAL  
FT CHAIN 17 >661 COMPLEMENT RECEPTOR 1.  
FT NON\_TER 661  
SQ SEQUENCE 661 AA; 72966 MW; 9D78E262 CRC32;

Query Match  
Best Local Similarity 100.0%; Score 422; DB 4; Length 661;  
Pred. No. 7 8e-40;  
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RIPCGLPTTNGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSDNDQV 60  
|||||  
Db 138 RIPCGLPTTNGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSDNDQV 197  
|||||

QY 61 GWSGAPACIIPNK 75  
|||||  
Db 198 GWSGAPACIIPNK 212  
|||||

RESULT 4  
029530 PRELIMINARY; PRT; 2014 AA.  
AC 029530;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE COMPLEMENT RECEPTOR 1 PRECURSOR (FRAGMENT).  
GN CRI.  
OS PAN TROGLODYTES (CHIMPANZEE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; PAN.  
LN 11  
RN SEQUENCE FROM N.A.  
RP MEDLINE; 94292799.  
RA BIRMINGHAM D.J., SHEN X.P., HOURCADE D., NICKELIS M.W., ATKINSON J.P.;  
RT "Primary sequence of an alternatively spliced form of CRI. Candidate  
RT for the 75,000 M(r) complement receptor expressed on chimpanzee  
erythrocytes";  
RL J. IMMUNOL. 153:691-700(1994).  
DR EMBL: L24921; G557727; .  
DR PRAM: PF00084; sush1; 9.  
KW SIGNAL; ALTERNATIVE SPLICING.  
FT NON\_TER 1  
FT SIGNAL 1  
FT CHAIN <1 16 POTENTIAL  
FT CHAIN 17 >661 COMPLEMENT RECEPTOR 1.  
FT NON\_TER 661  
SQ SEQUENCE 661 AA; 72966 MW; 9D78E262 CRC32;

Query Match	88.4%	Score 373	DB 4	Length 1911
Best Local Similarity	89.3%	Pred. No. 8	2e-34	
Matches	67	Conservative	2	Mismatches 6; Indels 0; Gaps 0
QY	1	RIPCLPTITNGDSDISTRENFHYGGSVYTYACNCGSGRKAFELVGEPSITCTSDDDY	60	
Db	500	RIPCLPTPIIAGDLSITRELFHYGGSVYTYACNCGSGRKAFELVGEPSITCTSDDDY	559	
QY	61	GIWSGPAPQCIIIPNK 75		
Db	560	GIWSGPAPQCIIIPNK 574		

	Query Match	85.1%	Score 359;	DB 4;	Length 522;	
	Best Local Similarity	87.8%	Pred.	No. 7-8e-33;		
Matches	65;	Conservative	3;	Mismatches	6;	Indels
					0;	Gaps
OY	2	IPCGLPPIITNGDDISTRTRENFHVGSVYYTKCNPSGSGRKRPVELGEFSIYTSDMDDOYG	61			
Dd	169	I PCGPPLP I A G D E L S R E F P T G S V Y T K C N L S G S G R K K L F E L V G E P S I Y T S K D O V G	228			
OY	62	I W S G P A P Q C I I P N K	75			
Dd	229	I W S G P A P Q C I I P N K	242			

8

```
ID 028797 PRELIMINARY; PRT; 522 AA.
AC 028797;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE UNKNOWN PROTEIN (FRAGMENT).
OS PAN TROGLODYTES (CHIMPANZEE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; PAN.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94292799.
RA BIRMINGHAM D.J., SHEN X.P., HOURCADE D., NICKELLS M.W., ATKINSON J.P.;
RT "Primary sequence of an alternatively spliced form of CRL. Candidate
RT for the 75,000 M(r) complement receptor expressed on chimpanzee
RT erythrocytes."
RL J. IMMUNOL. 153:691-700(1994).
DR EMBL; L24922; G557729; -.
DR PFM; PF00084; sushl; 8.
FT NON-TER
SO SEQUENCE 522 AA; 57304 MM; 21A32E12 CRC32;

Query Match 83.6%; Score 353; DB 4; Length 522;
Best Local Similarity 85.3%; Pred. No. 3.7e-32;
Matches 64; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

OY 1 RIPCGLPTINGDFTSIRENFHYSVYTRCNPBGSGRKVFELVGPSTICTSNDQV 60
DB 105 RIPCGLPTINGDFTSIRENFHYSVYTRCNPBGSGRKVFELVGPSTICTSKDDV 164
QY 61 GWSGPAPOCIIIPNK 75
DB 165 GWSGPAPOCIIIPNK 179

RESULT 9
OY 014079 PRELIMINARY; PRT; 479 AA.
AC 014079;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE COMPLEMENT RECEPTOR 1-LIKE PROTEIN (CRL) (FRAGMENT).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90110163.
RA HOURCADE D., MIESNER D.R., BEE C., ZELDES W., ATKINSON J.P.;
RT "Duplication and divergence of the amino-terminal coding region of the
RT complement receptor 1 (CRL) gene. An example of concerted (horizontal)
RT evolution within a gene."
RL J. BIOL. CHEM. 265:974-980(1990).
DR EMBL; M31231; G563326; JOINED.
DR EMBL; M31232; G563326; JOINED.
DR EMBL; M31233; G563326; JOINED.
DR EMBL; M31237; G563326; -.
DR EMBL; M31230; G563326; JOINED.
DR EMBL; M31234; G563326; JOINED.
DR EMBL; X14359; G563326; JOINED.
DR EMBL; X14360; G563326; JOINED.
DR EMBL; M31235; G563326; JOINED.
DR EMBL; M31236; G563326; JOINED.
DR PFM; PF00084; sushl; 7.
FT NON-TER
SO SEQUENCE 479 AA; 52533 MM; EC994E0D CRC32;

Query Match 83.2%; Score 351; DB 2; Length 479;
Best Local Similarity 85.3%; Pred. No. 5.7e-32;
Matches 64; Conservative 2; Mismatches 9; Indels 0; Gaps 0;
```

```
OY 1 RIPCGLPTINGDFTSIRENFHYSVYTRCNPBGSGRKVFELVGPSTICTSNDQV 60
DB 163 RIPCGLPTINGDFTSIRENFHYSVYTRCNPBGSGRKVFELVGPSTICTSKDDV 222
QY 61 GWSGPAPOCIIIPNK 75
DB 223 GWSGPAPOCIIIPNK 237

RESULT 10
OY 035520 PRELIMINARY; PRT; 417 AA.
AC 035520;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE 512 ANTIGEN (FRAGMENT).
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=LIVER.
RA DOHI N., SAKURADA C., NONAKA M., OKADA N., OKADA H.;
RT SUBMITTED (NOV-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; D42116; D1023417; -.
DR PFM; PF00084; sushl; 6.
FT NON-TER
SO SEQUENCE 417 AA; 45951 MM; A6F919B4 CRC32;

Query Match 69.2%; Score 292; DB 10; Length 417;
Best Local Similarity 68.9%; Pred. No. 2.2e-25;
Matches 51; Conservative 9; Mismatches 14; Indels 0; Gaps 0;

OY 2 IPCGLPTINGDFTSIRENFHYSVYTRCNPBGSGRKVFELVGPSTICTSNDQV 61
DB 21 IPCGLPTINGDFTSIRENFHYSVYTRCNPBGSGRKVFELVGPSTICTSIDQV 80
QY 62 IWSGPAPOCIIIPNK 75
DB 81 IWSGPAPOCIIIPNK 94

RESULT 11
OY 063612 PRELIMINARY; PRT; 497 AA.
AC 063612;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE 512 ANTIGEN PRECURSOR.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=LIVER.
RX MEDLINE; 94161746.
RA SAKURADA C., SENO H., DOHI N., TAKIZAWA H., NONAKA M., OKADA N.,
RA OKADA H.;
RT "Molecular cloning of the rat complement regulatory protein, 512
RT antigen."
RL BIOCHEM. BIOPHYS. RES. COMMUN. 198:819-826(1994).
DR EMBL; D42114; G987688; -.
DR PFM; PF00084; sushl; 6.
FT SIGNAL
FT CHAIN 1 36 POTENTIAL.
FT SIGNAL 37 497 512 ANTIGEN.
SO SEQUENCE 497 AA; 54786 MM; 076347C7 CRC32;
```

```

Query Match          69.2%; Score 292; DB 10; Length 497;
Best Local Similarity 68.9%; Pred. No. 2,7e-25;
Matches 51; Conservative 9; Mismatches 14; Indels 0; Gaps 0;

QY 2 IPGGLPTTNGDFISTNRENFHSGVVTYRCNPGSGGRKVFELVGEPSITCTSDNDQVG 61
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 159 IPCEIPSPINGDFESFNREDFFHGMVYTCNDTARGKFLNVLGEPSTHCTSDIGQVG 218
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 62 IWGAPQCIIIPNK 75
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 219 VWGSPPOCIETLKN 232

RESULT 12
ID 063135 PRELIMINARY; PRT: 559 AA.
AC 063135:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE COMPLEMENT REGULATORY PROTEIN.
GN CRRY.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY.
RX MEDLINE: 96006570.
RA OUTIG R.J., LO C.F., ALEXANDER J.J., SNEED A.E., MOXLEY G. III;
RT "Molecular characterization of rat Crry: widespread distribution of
   two alternative forms of Crry mRNA.";
RL IMMUNOGENETICS 42:362-367(1995).
RP SEQUENCE FROM N.A.
RC TISSUE-LIVER;
RA DOHI N., SAKURADA C., NONAKA M., OKADA N., OKADA H.;
RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: L36532; G1220295; -.
DR EMBL: D42115; D1023416; -.
DR PFAM: PF00084; sushi; 7.
SQ SEQUENCE 559 AA; 61680 MW; C87EC58 CRC32;

Query Match          69.2%; Score 292; DB 10; Length 559;
Best Local Similarity 68.9%; Pred. No. 3e-25;
Matches 51; Conservative 9; Mismatches 14; Indels 0; Gaps 0;

QY 2 IPGGLPTTNGDFISTNRENFHSGVVTYRCNPGSGGRKVFELVGEPSITCTSDNDQVG 61
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 159 IPCEIPSPINGDFESFNREDFFHGMVYTCNDTARGKFLNVLGEPSTHCTSDIGQVG 218
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 62 IWGAPQCIIIPNK 75
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 219 VWGSPPOCIETLKN 232

RESULT 13
ID 064735 PRELIMINARY; PRT: 483 AA.
AC 064735:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE COMPLEMENT RECEPTOR RELATED PROTEIN.
GN CRRY.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RX MEDLINE: 90171600.

```

```

RA PAUL M.S., AEGERTER-SHAW M., CEPER K., MILLER M.D., WEIS J.H.;
RT "The murine complement receptor gene family. III. The genomic and
   transpositional complexity of the Crry and Crry-ps genes.";
RL J. IMMUNOL. 144:1988-1996(1990).
DR EMBL: M34166; G387133; JOINED.
DR EMBL: M34167; G387133; JOINED.
DR EMBL: M34168; G387133; JOINED.
DR EMBL: M34169; G387133; JOINED.
DR EMBL: M34170; G387133; JOINED.
DR EMBL: M34171; G387133; JOINED.
DR EMBL: M34172; G387133; JOINED.
DR EMBL: M34173; G387133; JOINED.
DR EMBL: M34173; G387133; -.
DR EMBL: M34164; G387133; JOINED.
DR EMBL: M34165; G387133; JOINED.
DR EMBL: M34164; G387133; JOINED.
DR EMBL: M34170; G387133; JOINED.
DR EMBL: M34171; G387133; JOINED.
DR EMBL: M34165; G387133; JOINED.
DR EMBL: M34166; G387133; JOINED.
DR EMBL: M34167; G387133; JOINED.
DR EMBL: M34168; G387133; JOINED.
DR EMBL: M34169; G387133; JOINED.
DR MGD: MGI:88513; CRRY.
DR PFAM: PF00084; sushi; 5.
FT VARIANT 41 83 MISSING (IN LIVER ISOFORM).
SQ SEQUENCE 483 AA; 53762 MW; 70C73A80 CRC32;

Query Match          69.0%; Score 291; DB 10; Length 483;
Best Local Similarity 66.2%; Pred. No. 3.4e-25;
Matches 49; Conservative 10; Mismatches 15; Indels 0; Gaps 0;

QY 2 IPGGLPTTNGDFISTNRENFHSGVVTYRCNPGSGGRKVFELVGEPSITCTSDNDQVG 61
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 206 IPCEIPSPINGDFESFNREDFFHGMVYTCNDTARGKFLNVLGEPSTHCTSDIGQVG 265
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 62 IWGAPQCIIIPNK 75
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 266 VWGSPPOCIETLKN 279

RESULT 14
ID 099254 PRELIMINARY; PRT: 679 AA.
AC 099254:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE COMPLEMENT RECEPTOR TYPE 2 PRECURSOR (CR2)
   (ALTERNATIVELY SPLICED VERSION) (FRAGMENT).
GN CR2.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RX MEDLINE: 90229754.
RA KURTZ C.B., O'TOOLE E., CHRISTENSEN S.M., WEIS J.H.;
RT "The murine complement receptor gene family. IV. Alternative splicing
   of Cr2 gene transcripts predicts two distinct gene products that share
   homologous domains with both human CR2 and CRI.";
RL J. IMMUNOL. 144:3581-3591(1990).
RP SEQUENCE OF 21-367 FROM N.A.
RX MEDLINE: 95105691.
RA KIM Y.U., KINOSHITA T., MOLINA H., HOURCADE D., SEYA T., WAGNER L.M.,
RA HOLERS V.M.;
RT "Mouse complement regulatory protein Crry/p65 uses the specific
   mechanisms of both human decay-accelerating factor and membrane
   cofactor protein.";

```



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 00:42:53 ; Search time 58.51 seconds  
(without alignments)  
7.259 Million cell updates/sec

Title: US-09-142-043-4

Perfect score: 119  
Sequence: 1 CNPMSGGRKRVFELVGEPSIYC 21

Scoring table: BLOSUM62

Searched: 162890 segs, 20225328 residues

Database: A\_Geneseq\_34.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	119	100.0	2317	1 P92219	CRI protein New nu
2	119	100.0	2039	1 R11810	Human complement t
3	119	100.0	1537	1 R11982	Partial human comp
4	119	100.0	543	1 R28543	CRI-4 (35E, 37Y) a
5	119	100.0	543	1 R28544	CRI-4 (35E) analog
6	119	100.0	543	1 R28545	CRI-4 (37Y) analog
7	119	100.0	543	1 R28546	CRI-4 (44T, 47D, 4
8	119	100.0	543	1 R28547	CRI-4 (52S, 53S, 5
9	119	100.0	543	1 R28548	CRI-4 (57V, 59K) a
10	119	100.0	543	1 R28549	CRI-4 (64K, 65T) a
11	119	100.0	543	1 R28550	CRI-4 (64K) analog
12	119	100.0	543	1 R28551	CRI-4 (65T) analog
13	119	100.0	543	1 R28552	CRI-4 (78T, 79D) a
14	119	100.0	543	1 R28553	CRI-4 (85R, 87N) a
15	119	100.0	543	1 R28554	CRI-4 (92T, 94H) a
16	119	100.0	543	1 R28555	CRI-4 (92T) analog
17	119	100.0	543	1 R28556	CRI-4 (94H) analog
18	119	100.0	543	1 R28557	CRI-4 (99H, 103E)
19	119	100.0	543	1 R28558	CRI-4 (109N, 110A,
20	119	100.0	543	1 R28559	CRI-4 (114-117STK
21	119	100.0	543	1 R28560	CRI-4 (114S) analo
22	119	100.0	543	1 R28561	CRI-4 (115T) analo
23	119	100.0	543	1 R28562	CRI-4 (116K) analo
24	119	100.0	543	1 R28563	CRI-4 (117P) analo
25	119	100.0	543	1 R28564	CRI-4 (116K, 117P)
26	119	100.0	543	1 R28565	CRI-4 (121O) analo
27	119	100.0	483	1 R29091	CRI-4 (amino acids
28	119	100.0	481	1 R29092	CRI-4 (amino acids
29	119	100.0	543	1 R28566	CRI-4 (318R, 319N)
30	119	100.0	543	1 R28567	CRI-4 (318-321 RNP
31	119	100.0	543	1 R28568	CRI-4 (347T, 349Y)
32	119	100.0	543	1 R28569	CRI-4 (369-376 STK
33	119	100.0	543	1 R28570	CRI-4 (266-274 KTK
34	119	100.0	543	1 R28571	CRI-4 (364-367 NNA
35	119	100.0	2039	1 R36743	CRI. Nucleic acid
36	119	100.0	197	1 R47152	Sequence of solubl
37	119	100.0	76	1 R47153	Sequence of solubl
38	119	100.0	254	1 R47154	Sequence of solubl
39	119	100.0	133	1 R47155	Sequence of solubl
40	119	100.0	133	1 R47156	Sequence of solubl
41	119	100.0	21	1 W31835	Peptide 1 from the
42	119	100.0	211	1 W45909	SCR 1-3 of complem
43	119	100.0	209	1 W45912	SCR 1-3 of complem

## ALIGNMENTS

44 119 100.0 198 1 W45897 SCR 1-3 of complem  
45 119 100.0 1930 1 W45899 Human complement r

```
RESULT 1
P92219
ID P92219 standard; protein: 2317 AA.
AC P92219;
DT 22-FEB-1990 (first entry)
DE CRI protein
KW Complement; cofactor.
OS Homo sapiens (human).
FH Key Location/Qualifiers
FT peptide 10..50 /label= signal_peptide
FT
FT W08909220-A.
FT 05-OCT-1989.
FT 31-MAR-1989: U01358.
FT 01-APR-1988: US-176532.
FT (TCEU) T Cell Sciences Inc; (UYJO) The Johns Hopkins University;
FT (BRIG*) The Brigham and Women's Hospital.
FT Pearson DT, Klinkstein LB, Wong W, Carson G, Concino MF, Makrides SC;
FT MPI: 89-309498/42.
FT N-PSDB: N91477.
FT New nucleic acid sequences encoding new CRI protein - and its fragment,
FT for diagnosis and control of complement-related immune defects,
FT inflammation, myocardial infarct, etc
FT PS Claim 1, fig. 1; 191pp; English.
FT This is full-length CRI protein, and shortened forms are new, lacking
FT C4b, have cofactor activity and inhibit C3 and C5 convertase activity.
FT In the sequence, x-untranslated region. This has 7 short consensus
FT repeats (CR) of 4 long direct homologous repeats (LHR) (see fig. 10 in
FT patent). Deletion mutants, eg LHR-A for binding of C4b, and LHR-C
FT for C3b binding, be made, and LHR-B and -C can detect CR2 sequences.
FT CC They are useful in diagnosing and treating immune disorders, and prevent
FT CC perfusion injury.
FT SQ Sequence 2317 AA;

Query Match 100.0%; Score 119; DB 1; Length 2317;
Best Local Similarity 100.0%; Pred. No. 2.7e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPMSGGRKRVFELVGEPSIYC 21
DB 204 CNPMSGGRKRVFELVGEPSIYC 224

RESULT 2
R11810
ID R11810 standard; Protein: 2039 AA.
AC R11810;
DT 25-JUN-1991 (first entry)
DE Human complement type 1 receptor.
KW Complement system; C3b/C4b receptor; CRI; allergic reaction;
KW Immune response; clone lambda T109.1.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1..41 /label= putative signal peptide
FT 42..2039 /label= CRI
FT protein
FT W09105047-A.
FT 18-APR-1991.
FT 25-SEP-1990: U05454.
FT 26-SEP-1988: US-412745.
FT 26-SEP-1990: US-912349.
FT (TCEU) T CELL SCI INC.
FT (UYJO) JOHNS HOPKINS UNIVERSITY.
```







CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCRs of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-1 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-8. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQO  
CC accession number R11810 and descriptions in the disclosure.  
SO Sequence 543 AA:

Query Match 100.0%; Score 119; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 5,7e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CNPMSGGRKVFELVGEPSIYC 21  
DB 154 CNPMSGGRKVFELVGEPSIYC 174

RESULT 8  
R28547 ID R28547 standard; peptide: 543 AA.  
AC R28547;  
DT 19-MAR-1993 (first entry)  
DE CRI-4 (525, 535, 54P) analogue.  
KM short consensus repeat; regulator of complement activation;  
KW C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.  
FH key Location/Qualifiers  
FT region 1..60  
FT /label= SCR-1  
FT region 61..122  
FT /label= SCR-2  
FT region 451..510  
FT /label= SCR-8  
FT region 511..543  
FT /label= SCR-9  
FT /note= "TRUNCATED"  
FT misc-difference 52  
FT /note= "Thr substituted by Ser (SCR-8)"  
FT misc-difference 53  
FT /note= "Gly substituted by Ser (SCR-8)"  
FT misc-difference 54  
FT /note= "Ala substituted by Pro (SCR-8)"  
PN EP-512733-A.  
PD 11-NOV-1992.  
PF 28-APR-1992; 303826.  
PR 03-MAY-1991; US-695514.  
PA (UNITM) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Krych M;  
DR MPI: 92-375009/46.  
PT Complement activity regulator protein analogues - useful for  
PT treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11; Fig 2 and R11810; 23pp; English.  
CC The CDNA clone designated CRI-4 was described in J.Exp.Med.(1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCRs of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-1 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-8. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQO  
CC accession number R11810 and descriptions in the disclosure.  
SO Sequence 543 AA;

Query Match 100.0%; Score 119; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 5,7e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CNPMSGGRKVFELVGEPSIYC 21  
DB 154 CNPMSGGRKVFELVGEPSIYC 174

RESULT 9  
R28548 ID R28548 standard; peptide: 543 AA.  
AC R28548;  
DT 19-MAR-1993 (first entry)  
DE CRI-4 (57V, 59K) analogue.  
KM short consensus repeat; regulator of complement activation;  
KW C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.  
FH key Location/Qualifiers  
FT region 1..60  
FT /label= SCR-1  
FT region 61..122  
FT /label= SCR-2  
FT region 451..510  
FT /label= SCR-8  
FT region 511..543  
FT /label= SCR-9  
FT /note= "TRUNCATED"  
FT misc-difference 57  
FT /note= "Arg substituted by Val (SCR-8)"  
FT misc-difference 59  
FT /note= "Arg substituted by Lys (SCR-8)"  
PN EP-512733-A.  
PD 11-NOV-1992.  
PF 28-APR-1992; 303826.  
PR 03-MAY-1991; US-695514.  
PA (UNITM) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Krych M;  
DR MPI: 92-375009/46.  
PT Complement activity regulator protein analogues - useful for  
PT treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11; Fig 2 and R11810; 23pp; English.  
CC The CDNA clone designated CRI-4 was described in J.Exp.Med.(1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCRs of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-1 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-8. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQO  
CC accession number R11810 and descriptions in the disclosure.  
SO Sequence 543 AA;

Query Match 100.0%; Score 119; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 5,7e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CNPMSGGRKVFELVGEPSIYC 21  
DB 154 CNPMSGGRKVFELVGEPSIYC 174

RESULT 10  
R28549 ID R28549 standard; peptide: 543 AA.  
AC R28549;  
DT 19-MAR-1993 (first entry)

DE CRI-4 (64K, 65T) analogue.  
KW Short consensus repeat; regulator of complement activation;  
KW C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT region 1..60  
FT /label= SCR-1  
FT 61..122  
FT /label= SCR-2  
FT region 451..510  
FT /label= SCR-8  
FT 511..543  
FT /label= SCR-9  
FT /note= "TRUNCATED"  
FT misc\_difference 64  
FT /note= "Arg substituted by Lys (SCR-9)"  
FT misc\_difference 65  
FT /note= "Asn substituted by Thr (SCR-9)"  
PN EP-512733-A.  
PD 11-NOV-1992.  
PF 28-APR-1992; 303826.  
PR 03-MAY-1991; US-695514.  
PA (UNITW ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Krych M;  
DR WPI: 92-375009/46.  
PT Complement activity regulator protein analogues - useful for  
PT treating auto:immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11; Fig 2 and R1810: 23pp; English.  
CC The CDNA clone designated CRI-4 was described in J.Exp.Med.(1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCRs of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-2 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-9. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQ  
CC accession number R1810 and descriptions in the disclosure.  
CC Sequence 543 AA;  
SQ

Query Match 100.0%; Score 119; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 5,7e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPMSGGRKVFELVGEPSIYC 21  
DB 154 CNPMSGGRKVFELVGEPSIYC 174

RESULT 11  
R28550  
ID R28550 standard; peptide: 543 AA.  
AC R28550;  
DT 19-MAR-1993 (first entry)  
DE CRI-4 (64K) analogue.  
KW Short consensus repeat; regulator of complement activation;  
KW C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT region 1..60  
FT /label= SCR-1  
FT 61..122  
FT /label= SCR-2  
FT region 451..510  
FT /label= SCR-8  
FT 511..543  
FT /label= SCR-9  
FT /note= "TRUNCATED"  
FT misc\_difference 64

FT EP-512733-A. /note= "Arg substituted by Lys (SCR-9)"  
PN 11-NOV-1992.  
PD 28-APR-1992; 303826.  
PF 03-MAY-1991; US-695514.  
PR (UNITW ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Krych M;  
DR WPI: 92-375009/46.  
PT Complement activity regulator protein analogues - useful for  
PT treating auto:immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11; Fig 2 and R1810: 23pp; English.  
CC The CDNA clone designated CRI-4 was described in J.Exp.Med.(1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCRs of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-2 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-9. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQ  
CC accession number R1810 and descriptions in the disclosure.  
CC Sequence 543 AA;  
SQ

Query Match 100.0%; Score 119; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 5,7e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPMSGGRKVFELVGEPSIYC 21  
DB 154 CNPMSGGRKVFELVGEPSIYC 174

RESULT 12  
R28551  
ID R28551 standard; peptide: 543 AA.  
AC R28551;  
DT 19-MAR-1993 (first entry)  
DE CRI-4 (65T) analogue.  
KW Short consensus repeat; regulator of complement activation;  
KW C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT region 1..60  
FT /label= SCR-1  
FT 61..122  
FT /label= SCR-2  
FT region 451..510  
FT /label= SCR-8  
FT 511..543  
FT /label= SCR-9  
FT /note= "TRUNCATED"  
FT misc\_difference 64  
FT /note= "Arg substituted by Lys (SCR-9)"  
FT misc\_difference 65  
FT /note= "Asn substituted by Thr (SCR-9)"  
PN EP-512733-A.  
PD 11-NOV-1992.  
PF 28-APR-1992; 303826.  
PR 03-MAY-1991; US-695514.  
PA (UNITW ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Krych M;  
DR WPI: 92-375009/46.  
PT Complement activity regulator protein analogues - useful for  
PT treating auto:immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11; Fig 2 and R1810: 23pp; English.  
CC The CDNA clone designated CRI-4 was described in J.Exp.Med.(1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCRs of CRI. The invention concerns analogues of "regulator of

CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-2 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-9. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQ  
CC accession number R11810 and descriptions in the disclosure.  
SO Sequence 543 AA:

Query Match 100.0%; Score 119; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 5.7e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVFELVGPSTIC 21  
DB 154 CNPSSGGRKVFELVGPSTIC 174

## RESULT 13

ID R28552 standard; peptide: 543 AA.  
AC R28552;  
DT 19-MAR-1993 (first entry)  
DE CRI-4 (78T, 79D) analogue.  
KW Short consensus repeat; regulator of complement activation;  
KW C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.  
FH Key  
FT region Location/Qualifiers  
FT 1..60  
FT /label= SCR-1  
FT 61..122  
FT /label= SCR-2  
FT 451..510  
FT /label= SCR-8  
FT 511..543  
FT /label= SCR-9  
FT /note= "TRUNCATED"  
FT misc\_difference 78  
FT /note= "Lys substituted by Thr (SCR-9)"  
FT misc\_difference 79  
FT /note= "Gly substituted by Asp (SCR-9)"  
PN EP-512733-A.  
PD 11-NOV-1992.  
PF 28-APR-1992: 303826.  
PR 03-MAY-1991: US-695514.  
PA (UNIW ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Kirsch M;  
DR WPI: 92-375009/46.  
PT Complement activity regulator protein analogues - useful for  
PT treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11: Fig 2 and R11810; 23pp: English.  
CC The CDNA clone designated CRI-4 was described in J. Exp. Med. (1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCR-9 of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-2 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-9. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQ  
CC accession number R11810 and descriptions in the disclosure.  
SO Sequence 543 AA:

Query Match 100.0%; Score 119; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 5.7e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVFELVGPSTIC 21  
DB 154 CNPSSGGRKVFELVGPSTIC 174

## RESULT 14

ID R28553 standard; peptide: 543 AA.  
AC R28553;  
DT 19-MAR-1993 (first entry)  
DE CRI-4 (85R, 87N) analogue.  
KW Short consensus repeat; regulator of complement activation;  
KW C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.  
FH Key  
FT region Location/Qualifiers  
FT 1..60  
FT /label= SCR-1  
FT 61..122  
FT /label= SCR-2  
FT 451..510  
FT /label= SCR-8  
FT 511..543  
FT /label= SCR-9  
FT /note= "TRUNCATED"  
FT misc\_difference 85  
FT /note= "Gln substituted by Arg (SCR-9)"  
FT misc\_difference 87  
FT /note= "Lys substituted by Asn (SCR-9)"  
PN EP-512733-A.  
PD 11-NOV-1992.  
PF 28-APR-1992: 303826.  
PR 03-MAY-1991: US-695514.  
PA (UNIW ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Kirsch M;  
DR WPI: 92-375009/46.  
PT Complement activity regulator protein analogues - useful for  
PT treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11: Fig 2 and R11810; 23pp: English.  
CC The CDNA clone designated CRI-4 was described in J. Exp. Med. (1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCR-9 of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-2 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-9. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQ  
CC accession number R11810 and descriptions in the disclosure.  
SO Sequence 543 AA:

Query Match 100.0%; Score 119; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 5.7e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVFELVGPSTIC 21  
DB 154 CNPSSGGRKVFELVGPSTIC 174

## RESULT 15

ID R28554 standard; peptide: 543 AA.  
AC R28554;  
DT 19-MAR-1993 (first entry)  
DE CRI-4 (92T, 94H) analogue.  
KW Short consensus repeat; regulator of complement activation;  
KW C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.

```

FH Key Location/Qualifiers
FT region 1.60
FT /label= SCR-1
FT region 61.122
FT /label= SCR-2
FT region 451.510
FT /label= SCR-8
FT region 511.543
FT /label= SCR-9
FT /note= "TRUNCATED"
FT misc-difference 92
FT /note= "Lys substituted by Thr (SCR-9)"
FT misc-difference 94
FT /note= "Tyr substituted by His (SCR-9)"
FT EP-512733-A.
FT 11-NOV-1992.
FT 28-APR-1992; 303826
FT 03-MAY-1991; US-695514.
FT (UNIW ) UNIV WASHINGTON.
FT Atkinson JP, Hourcade D, Krych M;
FT WPI; 92-375009/46.
FT Complement activity regulator protein analogues - useful for
FT treating auto-immune diseases, to suppress transplant rejection,
FT for diagnosis etc.
FT Claim 11; Fig 2 and R1810; 23pp; English.
FT The cDNA clone designated CRI-4 was described in J. Exp. Med. (1988)
FT 168:1255-1270. It encodes the first 8 and a half amino terminal
FT SCRs of CRI. The invention concerns analogues of "regulator of
FT complement activation" proteins or truncated, hybrid or recombinant
FT forms of them. CRI-4 is a preferred truncated form and a number of
FT specified substitution variants of it are claimed in which certain
FT positions in SCR-2 which have been identified as important for the
FT degree of C3b- and C4b-binding are substituted by amino acids from
FT the corresponding positions in SCR-9. The specification does not
FT contain the CRI-4 sequence; the sequence given here was constructed
FT from the full-length CRI amino acid sequence having GENESEQ
FT accession number R1810 and descriptions in the disclosure.
FT Sequence 543 AA;
SO

```

```

Query Match 100.0%; Score 119; DB 1; Length 543;
Best Local Similarity 100.0%; Pred. No. 5.7e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CNPSSGGRKVFELVGPSTTC 21
DB 154 CNPSSGGRKVFELVGPSTTC 174

```

Search completed: July 18, 1999, 00:42:53  
Job time: 8586 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 06:07:56 ; Search time 39.54 Seconds  
(without alignments)  
5.241 Million cell updates/sec

Title: US-09-142-043-4  
Perfect score: 119  
Sequence: 1 CNPSSGGRKVFELVGPSPSYC 21

Scoring table: BLOSUM62  
Searched: 106577 seqs, 9868381 residues

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/iaa/ECTUS9\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/iaa/backfilest1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	119	100.0	197	2	US-08-356-361-27	Sequence 27, Appl
2	119	100.0	76	2	US-08-356-361-28	Sequence 28, Appl
3	119	100.0	254	2	US-08-356-361-29	Sequence 29, Appl
4	119	100.0	254	2	US-08-356-361-30	Sequence 30, Appl
5	119	100.0	133	2	US-08-356-361-31	Sequence 31, Appl
6	119	100.0	197	2	US-08-769-967A-27	Sequence 27, Appl
7	119	100.0	76	2	US-08-769-967A-28	Sequence 28, Appl
8	119	100.0	254	2	US-08-769-967A-29	Sequence 29, Appl
9	119	100.0	254	2	US-08-769-967A-30	Sequence 30, Appl
10	119	100.0	133	2	US-08-769-967A-31	Sequence 31, Appl
11	55	46.2	310	4	5256642-10	Patent No. 5256642
12	55	46.2	310	4	5472939-10	Patent No. 5472939
13	50.5	42.4	254	1	US-08-310-416A-13	Sequence 13, Appl
14	50.5	42.4	293	1	US-08-310-416A-16	Sequence 16, Appl
15	50.5	42.4	169	1	US-08-310-416A-18	Sequence 18, Appl
16	50.5	42.4	377	2	US-08-528-057-2	Sequence 2, Appl
17	50.5	42.4	370	2	US-08-528-057-42	Sequence 42, Appl
18	50.5	42.4	373	2	US-08-528-057-44	Sequence 44, Appl
19	50.5	42.4	324	2	US-08-528-057-46	Sequence 46, Appl
20	50.5	42.4	254	2	US-08-888-171-13	Sequence 13, Appl
21	50.5	42.4	293	2	US-08-888-171-16	Sequence 16, Appl
22	50.5	42.4	169	2	US-08-888-171-18	Sequence 18, Appl
23	50.5	42.4	254	2	US-08-435-149-1	Sequence 1, Appl
24	50.5	42.4	577	2	US-08-435-149-3	Sequence 3, Appl
25	50.5	42.0	324	1	US-08-310-416A-14	Sequence 14, Appl
26	50.5	42.0	323	2	US-08-888-171-14	Sequence 14, Appl
27	50.5	42.0	324	2	US-08-435-149-2	Sequence 2, Appl
28	50.5	42.0	260	4	5256642-5	Patent No. 5256642
29	50.5	42.0	260	4	5256642-6	Patent No. 5256642
30	50.5	42.0	260	4	5472939-5	Patent No. 5472939
31	50.5	42.0	250	4	5472939-6	Patent No. 5472939
32	49	41.2	10	4	5256642-18	Patent No. 5256642
33	49	41.2	10	4	5472939-18	Patent No. 5472939
34	46	38.7	1196	1	US-08-144-121-4	Sequence 4, Appl
35	45	37.8	1218	2	US-08-400-159-6	Sequence 6, Appl
36	42	35.3	190	1	US-08-044-621D-26	Sequence 26, Appl
37	42	35.3	190	1	US-08-044-621D-27	Sequence 27, Appl
38	42	35.3	190	1	US-08-044-621D-28	Sequence 28, Appl
39	42	35.3	190	2	US-08-709-912-14	Sequence 14, Appl

ALIGNMENTS

40	42	35.3	190	2	US-08-709-912-16	Sequence 16, Appl
41	42	35.3	190	2	US-08-709-912-17	Sequence 17, Appl
42	42	35.3	223	2	US-08-121-436A-2	Sequence 2, Appl
43	42	35.3	190	2	US-09-047-370-14	Sequence 14, Appl
44	42	35.3	190	2	US-09-047-370-16	Sequence 16, Appl
45	42	35.3	190	2	US-09-047-370-17	Sequence 17, Appl

RESULT 1  
US-08-356-361-27  
; Sequence 27, Application US/08356361  
; Patent No. 5833989  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Richard A.G.  
; APPLICANT: Dodd, Ian  
; APPLICANT: Freeman Mary A.  
; APPLICANT: Mossakowska, Danuta E.I.  
; TITLE OF INVENTION: No. 5833989el Compounds  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham-Corporate Intellectual Property  
; STREET: P.O. Box 1539  
; CITY: King of Prussia  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19406  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/356,361  
; FILING DATE: 03-Jul-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jervais, Herbert H.  
; REGISTRATION NUMBER: 31,171  
; REFERENCE/DOCKET NUMBER: P30423  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (610) 270-5019  
; TELEFAX: (610) 270-5090  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 197 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: N-terminal  
; US-08-356-361-27

Query Match 100.0% Score 119; DB 2; Length 197;  
Best Local Similarity 100.0%; Pred. No. 2.6e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CNPSSGGRKVFELVGPSPSYC 21  
DB 155 CNPSSGGRKVFELVGPSPSYC 175

RESULT 2  
US-08-356-361-28  
; Sequence 28, Application US/08356361  
; Patent No. 5833989  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Richard A.G.  
; APPLICANT: Dodd, Ian  
; APPLICANT: Freeman Mary A.  
; APPLICANT: Mossakowska, Danuta E.I.

TITLE OF INVENTION: NO. 5833989e1 Compounds  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham-Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,361  
FILING DATE: 03-Jul-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jervis, Herbert H.  
REGISTRATION NUMBER: 31,171  
REFERENCE/DOCKET NUMBER: P30423  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5019  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 76 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-356-361-28

Query Match 100.0%; Score 119; DB 2; Length 76;  
Best Local Similarity 100.0%; Pred. No. 9e-12;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVFELVGPSTYC 21  
Db 34 CNPSSGGRKVFELVGPSTYC 54

RESULT 3  
US-08-356-361-29  
Sequence 29, Application US/08356361  
Patent No. 5833989  
GENERAL INFORMATION:  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman Mary A.  
APPLICANT: Mossakowska, Danuta E.I.  
TITLE OF INVENTION: No. 5833989e1 Compounds  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham-Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,361  
FILING DATE: 03-Jul-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jervis, Herbert H.

REGISTRATION NUMBER: 31,171  
REFERENCE/DOCKET NUMBER: P30423  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5019  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-356-361-29

Query Match 100.0%; Score 119; DB 2; Length 254;  
Best Local Similarity 100.0%; Pred. No. 3.4e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVFELVGPSTYC 21  
Db 155 CNPSSGGRKVFELVGPSTYC 175

RESULT 4  
US-08-356-361-30  
Sequence 30, Application US/08356361  
Patent No. 5833989  
GENERAL INFORMATION:  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman Mary A.  
APPLICANT: Mossakowska, Danuta E.I.  
TITLE OF INVENTION: No. 5833989e1 Compounds  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham-Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,361  
FILING DATE: 03-Jul-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jervis, Herbert H.  
REGISTRATION NUMBER: 31,171  
REFERENCE/DOCKET NUMBER: P30423  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5019  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-356-361-30

Query Match 100.0%; Score 119; DB 2; Length 254;  
Best Local Similarity 100.0%; Pred. No. 3.4e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVFELVGPSTYC 21



Db 155 CNPGSGGRKVFELVGEPSIYC 175

## RESULT 5

US-08-356-361-31  
Sequence 31, Application US/08356361  
Patent No. 5833989  
GENERAL INFORMATION:  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman Mary A.  
APPLICANT: Mossakowska, Danuta E.I.  
TITLE OF INVENTION: No. 5833989el Compounds  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham-Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,361  
FILING DATE: 03-Jul-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jervis, Herbert H.  
REGISTRATION NUMBER: 31,171  
REFERENCE/DOCKET NUMBER: P30423  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5090  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 133 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-356-361-31

Query Match 100.0%; Score 119; DB 2; Length 133;  
Best local Similarity 100.0%; Pred. No. 1.7e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPGSGGRKVFELVGEPSIYC 21  
Db 34 CNPGSGGRKVFELVGEPSIYC 54

## RESULT 6

US-08-769-967A-27  
Sequence 27, Application US/08769967A  
Patent No. 5859223  
GENERAL INFORMATION:  
APPLICANT: Mossakowska, Danuta E.I.  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman, Anne Mary  
TITLE OF INVENTION: Soluble CRI Derivatives  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania

COUNTRY: USA  
ZIP: 19406

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,967A  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,569  
FILING DATE: 15-May-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: P30423C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5364  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 197 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-769-967A-27

Query Match 100.0%; Score 119; DB 2; Length 197;  
Best local Similarity 100.0%; Pred. No. 2.6e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPGSGGRKVFELVGEPSIYC 21  
Db 155 CNPGSGGRKVFELVGEPSIYC 175

## RESULT 7

US-08-769-967A-28  
Sequence 28, Application US/08769967A  
Patent No. 5859223  
GENERAL INFORMATION:  
APPLICANT: Mossakowska, Danuta E.I.  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman, Anne Mary  
TITLE OF INVENTION: Soluble CRI Derivatives  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,967A  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,569  
FILING DATE: 15-May-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954

REFERENCE/DOCKET NUMBER: P30423C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5364  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 76 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-769-967A-28

Query Match 100.0%; Score 119; DB 2; Length 76;  
Best Local Similarity 100.0%; Pred. No. 9e-12;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CNPSSGGRKVFELVGEPSIYC 21  
|||||  
Db 34 CNPSSGGRKVFELVGEPSIYC 54

RESULT 8  
US-08-769-967A-29  
Sequence 29, Application US/08769967A  
Patent No. 5859223  
GENERAL INFORMATION:  
APPLICANT: Mossakowska, Danuta E.I.  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman, Anne Mary  
TITLE OF INVENTION: Soluble CRI Derivatives  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Smithline Beecham Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,967A  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,569  
FILING DATE: 15-May-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: P30423C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5364  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-769-967A-29

Query Match 100.0%; Score 119; DB 2; Length 254;  
Best Local Similarity 100.0%; Pred. No. 3.4e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CNPSSGGRKVFELVGEPSIYC 21  
|||||  
Db 155 CNPSSGGRKVFELVGEPSIYC 175

RESULT 9  
US-08-769-967A-30  
Sequence 30, Application US/08769967A  
Patent No. 5859223  
GENERAL INFORMATION:  
APPLICANT: Mossakowska, Danuta E.I.  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman, Anne Mary  
TITLE OF INVENTION: Soluble CRI Derivatives  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Smithline Beecham Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,967A  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,569  
FILING DATE: 15-May-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: P30423C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5364  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-769-967A-30

Query Match 100.0%; Score 119; DB 2; Length 254;  
Best Local Similarity 100.0%; Pred. No. 3.4e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CNPSSGGRKVFELVGEPSIYC 21  
|||||  
Db 155 CNPSSGGRKVFELVGEPSIYC 175

RESULT 10  
US-08-769-967A-31  
Sequence 31, Application US/08769967A  
Patent No. 5859223  
GENERAL INFORMATION:  
APPLICANT: Mossakowska, Danuta E.I.  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman, Anne Mary  
TITLE OF INVENTION: Soluble CRI Derivatives  
NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: SmithKline Beecham Corporate Intellectual Property  
 STREET: P.O. Box 1539  
 CITY: King of Prussia  
 STATE: Pennsylvania  
 COUNTRY: USA  
 ZIP: 19406  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/769,967A  
 FILING DATE:  
 CLASSIFICATION: 536  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/440,569  
 FILING DATE: 15-May-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: King, William T.  
 REGISTRATION NUMBER: 30,954  
 REFERENCE/DOCKET NUMBER: P30423C2  
 TELEPHONE: (610) 270-5364  
 TELEFAX: (610) 270-5090  
 INFORMATION FOR SEQ ID NO: 31:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 133 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 FRAGMENT TYPE: internal  
 US-08-769-967A-31

Query Match 100.0%; Score 119; DB 2; Length 133;  
 Best Local Similarity 100.0%; Pred. No. 1,7e-11;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CNPSSGGRKVELVGEPSIYC 21  
 DB 34 CNPSSGGRKVELVGEPSIYC 54

RESULT 11  
 5256642-10  
 Patent No. 5256642  
 APPLICANT: FEARON, DOUGLAS T.; KLICKSTEIN, LLOYD B.; WONG,  
 WINNIE W.; CARSON, GERALD R.; CONCINO, MICHAEL F.; IP, STEPHEN  
 H.; MAKRIDES, SAVVAS; MARSH, HENRY C. JR.  
 TITLE OF INVENTION: COMPOSITIONS OF SOLUBLE COMPLEMENT  
 RECEPTOR 1 (CR1) AND A THROMBOCYTIC AGENT, AND THE METHODS OF  
 USE THEREOF  
 NUMBER OF SEQUENCES: 30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/588,128  
 FILING DATE: 24-SEP-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 412,745  
 FILING DATE: 26-SEP-1989  
 APPLICATION NUMBER: 332,865  
 FILING DATE: 03-APR-1989  
 APPLICATION NUMBER: 176,532  
 FILING DATE: 01-APR-1988  
 SEQ ID NO: 10:  
 LENGTH: 1847  
 5256642-10

Query Match 46.2%; Score 55; DB 4; Length 310;  
 Best Local Similarity 100.0%; Pred. No. 0.3;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ELVGEPSIYC 21  
 DB 111 ELVGEPSIYC 120

RESULT 12  
 5472939-10  
 Patent No. 5472939  
 APPLICANT: FEARON, DOUGLAS T.; KLICKSTEIN, LLOYD B.; WONG,  
 WINNIE W.; CARSON, GERALD R.; CONCINO, MICHAEL F.; IP, STEPHEN  
 H.; MAKRIDES, SAVVAS; MARSH, HENRY C. JR.  
 TITLE OF INVENTION: METHOD OF TREATING COMPLEMENT  
 MEDIATED DISORDERS  
 NUMBER OF SEQUENCES: 30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/138,825  
 FILING DATE: 19-OCT-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 588,128  
 FILING DATE: 24-SEP-1990  
 APPLICATION NUMBER: 412,745  
 FILING DATE: 26-SEP-1989  
 APPLICATION NUMBER: 332,865  
 FILING DATE: 03-APR-1989  
 APPLICATION NUMBER: 176,532  
 FILING DATE: 01-APR-1988  
 SEQ ID NO: 10:  
 LENGTH: 2006  
 5472939-10

Query Match 46.2%; Score 55; DB 4; Length 310;  
 Best Local Similarity 100.0%; Pred. No. 0.3;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ELVGEPSIYC 21  
 DB 111 ELVGEPSIYC 120

RESULT 13  
 US-08-310-416A-13  
 Sequence 13, Application US/08310416A  
 Patent No. 5679546  
 GENERAL INFORMATION:  
 APPLICANT: Jone-Long Ko et al.  
 TITLE OF INVENTION: CHIMERIC PROTEINS WHICH BLOCK  
 TITLE OF INVENTION: COMPLEMENT ACTIVATION  
 NUMBER OF SEQUENCES: 19  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Fish & Richardson P.C.  
 STREET: 225 Franklin Street  
 CITY: Boston  
 STATE: Massachusetts  
 COUNTRY: U.S.A.  
 ZIP: 02110-2804  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
 COMPUTER: IBM PS/2 Model 502 or 55SX  
 OPERATING SYSTEM: MS-DOS (Version 5.0)  
 SOFTWARE: WordPerfect (Version 5.1)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/310,416A  
 FILING DATE: 22-SEP-1994  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Paul T. Clark  
 REGISTRATION NUMBER: 30,162  
 REFERENCE/DOCKET NUMBER: 06180/005001  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 542-5070  
 TELEFAX: (617) 542-8906



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 06:07:03 ; Search time 44.47 seconds  
(without alignments)  
17.690 Million cell updates/sec

Title: US-09-142-043-4

Perfect score: 119  
Sequence: 1 CNPSSGGRKVFELVGEPSIYC 21

Scoring table: BLOSUM62

Searched: 116738 seqs, 37460341 residues

Database: PIR\_58:\*\*

1: pir1:\*\*  
2: pir2:\*\*  
3: pir3:\*\*  
4: pir4:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	119	100.0	2014	2	I36936	complement recepto
2	119	100.0	661	2	I36937	complement recepto
3	119	100.0	2489	2	I73012	complement C3b/C4b
4	119	100.0	2039	2	A28507	complement C3b/C4b
5	98	82.4	482	2	A34924	complement C3b/C4b
6	77	64.7	497	2	JC2054	complement regulat
7	74	62.2	433	2	A30550	complement C3b/C4b
8	74	62.2	440	2	A43519	complement recepto
9	69	58.0	676	2	A45900	complement C3b rec
10	52	43.7	1091	1	PL0009	complement C3d/Eps
11	50.5	42.4	369	2	I57998	membrane cofactor
12	50.5	42.4	384	2	S01896	membrane cofactor
13	50.5	42.4	377	2	I54479	membrane cofactor
14	50.5	42.4	349	2	G02913	sperm CD46 - human
15	50	42.0	381	1	B26359	decay-accelerating
16	50	42.0	440	2	A26359	decay-accelerating
17	50	42.0	611	2	F70325	conserved hypotet
18	49	41.2	1025	2	A43526	complement C3d/Eps
19	48.5	40.8	369	2	JC5138	membrane cofactor
20	48.5	40.8	362	2	JC5194	membrane cofactor
21	47	39.5	340	2	I56234	decay-accelerating
22	46	38.7	1786	1	MMHUB1	laminin beta-1 cha
23	46	38.7	1786	1	MMHUB1	laminin beta-1 cha
24	45	37.8	469	1	NBMSC4	C4b-binding protei
25	45	37.8	217	2	B64693	D-ribulose-5-phosp
26	44.5	37.4	477	1	DMDXAF	2-hydroxyglutaryl-
27	44	37.0	579	2	A56740	sperm-egg recognit
28	44	37.0	1256	2	JE0209	brain-specific and
29	43.5	36.6	321	2	S62015	SOL protein - yea
30	43.5	36.6	156	2	A48886	8-oxo-7,8-dihydrog
31	43	36.1	597	2	S53711	C4BP alpha chain p
32	43	36.1	274	2	S22308	type II site-speci
33	43	36.1	1146	2	S59376	hypothetical prote
34	43	36.1	1146	2	S59376	hypothetical prote
35	43	36.1	370	2	C71041	hypothetical prote
36	42.5	35.7	1747	2	A54121	collagen alpha-4 c
37	42	35.3	190	1	A44593	endo-1,4-beta-xyla
38	42	35.3	190	1	A44595	endo-1,4-beta-xyla
39	42	35.3	190	1	A44594	endo-1,4-beta-xyla

40 42 35.3 165 2 I41043 dihydrofolate redu  
41 42 35.3 165 2 S60665 dihydrofolate redu  
42 42 35.3 159 2 S32014 dihydrofolate redu  
43 42 35.3 222 2 S39154 xylanase 1 - fungu  
44 42 35.3 223 2 S39883 endo-1,4-beta-xyla  
45 42 35.3 662 2 D40228 neurixin II-beta p

## ALIGNMENTS

RESULT 1  
I36936  
complement receptor 1 - chimpanzee (fragment)  
C:Species: Pan troglodytes (chimpanzee)  
C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 07-Aug-1998  
C:Accession: I36936  
R:Birmingham, D.J.; Shen, X.P.; Hourcade, D.; Nickells, M.W.; Atkinson, J.P.  
J. Immunol. 153, 691-700, 1994  
A>Title: Primary sequence of an alternatively spliced form of CRI. Candidate for the  
A:Reference number: I36935; MUID:94292799  
A:Accession: I36936  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-2014 <RES>  
A:Cross-references: GB:I24920; NID:9551564; PID:9557725  
C:Superfamily: complement factor H repeat homology  
F:18-74/Domain: complement factor H repeat homology <FH01>  
F:79-136/Domain: complement factor H repeat homology <FH02>  
F:141-207/Domain: complement factor H repeat homology <FH03>  
F:333-391/Domain: complement factor H repeat homology <FH12>  
F:1041-1107/Domain: complement factor H repeat homology <FH26>  
F:1749-1815/Domain: complement factor H repeat homology <FH27>

Query Match 100.0%; Score 119; DB 2; Length 2014;  
Best Local Similarity 100.0%; Pred. No. 8e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CNPSSGGRKVFELVGEPSIYC 21  
DB 170 CNPSSGGRKVFELVGEPSIYC 190

RESULT 2  
I36937  
complement receptor 1 - chimpanzee (fragment)  
C:Species: Pan troglodytes (chimpanzee)  
C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 07-Aug-1998  
C:Accession: I36937  
R:Birmingham, D.J.; Shen, X.P.; Hourcade, D.; Nickells, M.W.; Atkinson, J.P.  
J. Immunol. 153, 691-700, 1994  
A>Title: Primary sequence of an alternatively spliced form of CRI. Candidate for the  
A:Reference number: I36935; MUID:94292799  
A:Accession: I36937  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-661 <RES>  
A:Cross-references: GB:I24921; NID:9557726; PID:9557727  
C:Superfamily: complement factor H repeat homology  
F:18-74/Domain: complement factor H repeat homology <FH01>  
F:79-136/Domain: complement factor H repeat homology <FH02>  
F:141-207/Domain: complement factor H repeat homology <FH03>  
F:213-268/Domain: complement factor H repeat homology <FH04>  
F:333-391/Domain: complement factor H repeat homology <FH22>  
F:396-462/Domain: complement factor H repeat homology <FH27>

Query Match 100.0%; Score 119; DB 2; Length 661;  
Best Local Similarity 100.0%; Pred. No. 2.7e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CNPSSGGRKVFELVGEPSIYC 21

Db 170 CNPSSGGRKVELVEPSIYC 190

## RESULT 3

173012  
Complement C3b/C4b receptor (allotype S) precursor - human  
N:Alternate names: complement receptor type 1 (CRI); surface glycoprotein CD35  
C:Species: Homo sapiens (man)  
C:Date: 02-Jul-1996 #sequence, revision 02-Jul-1996 #text, change 24-Sep-1998  
C:Accession: 173012; A47602; S03291  
R:VLK, D.P.; Wong, W.W.  
J. Immunol. 151, 6214-6224, 1993  
A:Title: Structure of the gene for the F allele of complement receptor type 1 and sequen  
A:Reference number: 156203; MUID:94065175  
A:Accession: 173012  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2489 <RES>  
A:Cross-references: GB:117418; NID:9306678; PID:9451303  
R:Mong, W.W.; Cahill, J.M.; Rosen, M.D.; Kennedy, C.A.; Bonaccio, E.T.; Morris, M.J.; W  
J. Exp. Med. 169, 847-863, 1989  
A:Title: Structure of the human CRI gene. Molecular basis of the structural and quantita  
A:Reference number: A47602  
A:Accession: A47602  
A:Molecule type: DNA  
A:Residues: 1-41 <WON>  
A:Cross-references: GB:X14893  
R:Hourcade, D.; Miesner, D.R.; Atkinson, J.P.; Holers, V.M.  
J. Exp. Med. 168, 1255-1270, 1988  
A:Title: Identification of an alternative polyadenylation site in the human C3b/C4b rece  
type 1.  
A:Reference number: S03291; MUID:89010527  
A:Accession: S03291  
A:Molecule type: mRNA  
A:Residues: 26-584 <HOU>  
A:Cross-references: EMBL:X14362; NID:930197; PID:9736240  
A:Experimental source: clone CRI-4  
C:Genetics:  
A:Gene: GDB:CR1; CD35  
A:Cross-references: GDB:119800; OMIM:120620  
A:Map position: 1q32-q32  
A:Intons: 41/1; 101/1; 134/2; 163/1; 296/1; 355/1; 389/2; 418/1; 491/1; 551/1; 584/2; 6  
; 1484/2; 1513/1; 1646/1; 1705/1; 1739/2; 1768/1; 1844/1; 1904/1; 1937/2; 1966/1; 2099/1  
C:Keywords: duplication; glycoprotein; transmembrane protein  
F:1-41/Domain: signal sequence #status predicted <SIG>  
F:42-584/Product: complement C3b/C4b receptor, secreted #status predicted <MAT>  
F:43-99/Domain: complement factor H repeat homology <FH01>  
F:104-161/Domain: complement factor H repeat homology <FH02>  
F:166-232/Domain: complement factor H repeat homology <FH03>  
F:238-293/Domain: complement factor H repeat homology <FH04>  
F:297-353/Domain: complement factor H repeat homology <FH05>  
F:356-416/Domain: complement factor H repeat homology <FH06>  
F:421-467/Domain: complement factor H repeat homology <FH07>  
F:554-611/Domain: complement factor H repeat homology <FH08>  
F:616-682/Domain: complement factor H repeat homology <FH09>  
F:808-866/Domain: complement factor H repeat homology <FHXB>  
F:1004-1061/Domain: complement factor H repeat homology <FH09>  
F:1066-1132/Domain: complement factor H repeat homology <FH10>  
F:1138-1193/Domain: complement factor H repeat homology <FH11>  
F:1197-1253/Domain: complement factor H repeat homology <FH12>  
F:1258-1316/Domain: complement factor H repeat homology <FH13>  
F:1321-1387/Domain: complement factor H repeat homology <FH14>  
F:1393-1449/Domain: complement factor H repeat homology <FH15>  
F:1454-1511/Domain: complement factor H repeat homology <FH16>  
F:1516-1582/Domain: complement factor H repeat homology <FH17>  
F:1588-1643/Domain: complement factor H repeat homology <FH18>  
F:1647-1703/Domain: complement factor H repeat homology <FH19>  
F:1708-1766/Domain: complement factor H repeat homology <FH20>  
F:1771-1837/Domain: complement factor H repeat homology <FH21>  
F:1846-1902/Domain: complement factor H repeat homology <FH22>  
F:1907-1964/Domain: complement factor H repeat homology <FH23>

F:1969-2035/Domain: complement factor H repeat homology <FHXC>  
F:2100-2156/Domain: complement factor H repeat homology <FH24>  
F:2161-2219/Domain: complement factor H repeat homology <FHXD>

Query Match 100.0%; Score 119; DB 2; Length 2489;  
Best Local Similarity 100.0%; Pred. No. 9, 9e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 195 CNPSSGGRKVELVEPSIYC 215

## RESULT 4

A28507  
Complement C3b/C4b receptor precursor, membrane-bound (allotype F) - human  
N:Alternate names: complement receptor type 1 (CRI); surface glycoprotein CD35  
C:Species: Homo sapiens (man)  
C:Date: 19-Nov-1988 #sequence, revision 06-Sep-1996 #text, change 10-Sep-1997  
C:Accession: S03843; A28507; 156203; A24748; B24748; C24748  
R:Klickstein, L.B.; Barlow, T.J.; Miletic, V.; Rabson, L.D.; Smith, J.A.; Fearon, D.T  
J. Exp. Med. 168, 1699-1717, 1988  
A:Title: Identification of distinct C3b and C4b recognition sites in the human C3b/C4  
A:Reference number: S03843; MUID:89035992  
A:Accession: S03843  
A:Status: translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-2039 <KL>  
A:Cross-references: EMBL:Y00816; NID:930185; PID:930186  
R:Klickstein, L.B.; Wong, W.W.; Smith, J.A.; Weis, J.H.; Wilson, J.G.; Fearon, D.T.  
J. Exp. Med. 165, 1095-1112, 1987  
A:Title: Human C3b/C4b receptor (CRI). Demonstration of long homologous repeating dom  
A:Reference number: A28507; MUID:87168191  
A:Accession: A28507  
A:Molecule type: mRNA  
A:Residues: 503-771, 'FV', 774-2039 <KL2>  
R:VLK, D.P.; Wong, W.W.  
J. Immunol. 151, 6214-6224, 1993  
A:Title: Structure of the gene for the F allele of complement receptor type 1 and seq  
A:Reference number: 156203; MUID:94065175  
A:Accession: 156203  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-683, 'X', 685-1021, 'V', 1023-1614, 'V', 1616-1826, 'R', 1828-1849, 'D', 1851-187  
A:Cross-references: GB:117418; NID:9306678; PID:9306680  
R:Mong, W.W.; Klickstein, L.B.; Smith, J.A.; Weis, J.H.; Fearon, D.T.  
Proc. Natl. Acad. Sci. U.S.A. 82, 7711-7715, 1985  
A:Title: Identification of a partial cDNA clone for the human receptor for complement  
A:Reference number: A94073; MUID:86067975  
A:Accession: A24748  
A:Molecule type: mRNA  
A:Residues: 311-333; 729-745; 831-845 <WON>  
C:Genetics:  
A:Gene: GDB:CR1; CD35  
A:Cross-references: GDB:119800; OMIM:120620  
A:Map position: 1q32-q32  
A:Intons: 41/1; 101/1; 134/2; 163/1; 296/1; 355/1; 389/2; 418/1; 491/1; 551/1; 584/2  
; 1487/2; 1516/1; 1649/1; 1708/1; 1742/2; 1771/1; 1847/1; 1906/1; 1968/1; 1976/1; 200  
C:Keywords: duplication; glycoprotein; transmembrane protein  
F:1-41/Domain: signal sequence #status predicted <SIG>  
F:42-2039/Product: complement C3b/C4b receptor, membrane-bound #status predicted <MAT  
F:43-99/Domain: complement factor H repeat homology <FH01>  
F:104-161/Domain: complement factor H repeat homology <FH02>  
F:166-232/Domain: complement factor H repeat homology <FH03>  
F:238-293/Domain: complement factor H repeat homology <FH04>  
F:297-353/Domain: complement factor H repeat homology <FH05>  
F:358-416/Domain: complement factor H repeat homology <FH06>  
F:421-467/Domain: complement factor H repeat homology <FH07>  
F:493-549/Domain: complement factor H repeat homology <FH08>  
F:554-611/Domain: complement factor H repeat homology <FH09>

F:688-743/Domain: complement factor H repeat homology <FH10>  
F:747-803/Domain: complement factor H repeat homology <FH11>  
F:808-866/Domain: complement factor H repeat homology <FH12>  
F:871-937/Domain: complement factor H repeat homology <FH13>  
F:943-999/Domain: complement factor H repeat homology <FH14>  
F:1004-1061/Domain: complement factor H repeat homology <FH15>  
F:1066-1132/Domain: complement factor H repeat homology <FH16>  
F:1138-1193/Domain: complement factor H repeat homology <FH17>  
F:1197-1253/Domain: complement factor H repeat homology <FH18>  
F:1258-1316/Domain: complement factor H repeat homology <FH19>  
F:1321-1387/Domain: complement factor H repeat homology <FH20>  
F:1396-1454/Domain: complement factor H repeat homology <FH21>  
F:1457-1514/Domain: complement factor H repeat homology <FH22>  
F:1519-1585/Domain: complement factor H repeat homology <FH23>  
F:1591-1646/Domain: complement factor H repeat homology <FH24>  
F:1650-1706/Domain: complement factor H repeat homology <FH25>  
F:1711-1769/Domain: complement factor H repeat homology <FH26>  
F:1774-1840/Domain: complement factor H repeat homology <FH27>  
F:1848-1904/Domain: complement factor H repeat homology <FH28>  
F:1909-1965/Domain: complement factor H repeat homology <FH29>

Query Match 100.0%; Score 119; DB 2; Length 2039;  
Best Local Similarity 100.0%; Pred. No. 8.1e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CNPSSGGRKVFELVGPSTYC 21  
DB 195 CNPSSGGRKVFELVGPSTYC 215

RESULT 5  
A34924  
Complement C3b/C4b receptor-like protein precursor - human

C:Species: Homo sapiens (man)  
C:Date: 27-Jul-1990 #sequence\_revision 11-Apr-1997 #text\_change 11-Apr-1997  
C:Accession: A34924; S03292  
R:Hourcade, D.; Miesner, D.R.; Bee, C.; Zeldes, W.; Atkinson, J.P.  
J. Biol. Chem. 265, 974-980, 1990  
A:Title: Duplication and divergence of the amino-terminal coding region of the complement  
A:Reference number: A34924; MUID:90110163  
A:Accession: A34924  
A:Molecule type: DNA  
A:Residues: 1-479 <HOU>  
A:Cross-references: GB:J05195  
R:Hourcade, D.; Miesner, D.R.; Atkinson, J.P.; Holers, V.M.  
J. Exp. Med. 168, 1255-1270, 1988  
A:Title: Identification of an alternative polyadenylation site in the human C3b/C4b receptor  
type 1.  
A:Reference number: S03291; MUID:89010527  
A:Accession: S03292  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: FPFPAFR, 452-482 <HO2>  
A:Cross-references: EMBL:X14360  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1988

C:Superfamily: complement factor H repeat homology  
F:43-99/Domain: complement factor H repeat homology <FH01>  
F:104-161/Domain: complement factor H repeat homology <FH02>  
F:166-232/Domain: complement factor H repeat homology <FH03>  
F:238-293/Domain: complement factor H repeat homology <FH04>  
F:297-353/Domain: complement factor H repeat homology <FH05>  
F:358-416/Domain: complement factor H repeat homology <FH06>  
F:421-478/Domain: complement factor H repeat homology <FH07>

Query Match 82.4%; Score 98; DB 2; Length 482;  
Best Local Similarity 85.7%; Pred. No. 2.9e-07;  
Matches 18; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CNPSSGGRKVFELVGPSTYC 21  
DB 195 CNPSSGGRKVFELVGPSTYC 215

RESULT 6  
JC2054  
Complement regulatory protein, 512 antigen precursor - rat

C:Species: Rattus norvegicus (Norway rat)  
C:Date: 27-Jun-1994 #sequence\_revision 27-Jun-1994 #text\_change 22-Nov-1996  
C:Accession: JC2054; PC2027  
R:Sakurada, C.; Seno, H.; Dohi, N.; Takizawa, H.; Nonaka, M.; Okada, N.; Okada, H.  
Biochem. Biophys. Res. Commun. 198, 819-826, 1994  
A:Title: Molecular cloning of the rat complement regulatory protein, 512 antigen.  
A:Reference number: JC2054  
A:Accession: JC2054  
A:Molecule type: mRNA  
A:Residues: 1-497 <SAK>  
A:Accession: PC2027  
A:Molecule type: protein

A:Residues: 39-51 <SA2>  
C:Comment: This protein plays a critical role in protection against complement mediated  
C:Superfamily: complement factor H repeat homology  
C:Keywords: glycoprotein; transmembrane protein  
F:1-36/Domain: signal sequence #status predicted <SIG>  
F:37-488/Product: complement regulatory protein, 512 antigen #status predicted <MAR>  
F:38-94/Domain: complement factor H repeat homology <FH1>  
F:99-156/Domain: complement factor H repeat homology <FH2>  
F:161-227/Domain: complement factor H repeat homology <FH3>  
F:233-288/Domain: complement factor H repeat homology <FH4>  
F:294-353/Domain: complement factor H repeat homology <FH5>  
F:357-413/Domain: complement factor H repeat homology <FH6>  
F:421-444/Domain: transmembrane #status predicted <TM>  
F:247,331,346,450,482,483/Binding site: carbohydrate (asn) (covalent) #status predict

Query Match 64.7%; Score 77; DB 2; Length 497;  
Best Local Similarity 61.9%; Pred. No. 0.00044;  
Matches 13; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 CNPSSGGRKVFELVGPSTYC 21  
DB 190 CMTDARGKLFNLVGPSTYC 210

RESULT 7  
A30550

Complement C3b/C4b receptor precursor - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 03-Mar-1989 #sequence\_revision 03-Mar-1989 #text\_change 12-May-1995  
C:Accession: A30550  
R:Paul, M.S.; Aegerter, M.; O'Brien, S.E.; Kurtz, C.B.; Weis, J.H.  
J. Immunol. 142, 582-589, 1989  
A:Title: The murine complement receptor gene family. Analysis of mcr1 gene products a  
A:Reference number: A30550; MUID:89093944  
A:Accession: A30550  
A:Status: Preliminary  
A:Molecule type: mRNA  
A:Residues: 1-433 <PAU>  
C:Superfamily: complement factor H repeat homology  
F:42-98/Domain: complement factor H repeat homology <FH1>  
F:103-160/Domain: complement factor H repeat homology <FH2>  
F:165-231/Domain: complement factor H repeat homology <FH3>  
F:237-293/Domain: complement factor H repeat homology <FH4>  
F:299-355/Domain: complement factor H repeat homology <FH5>

Query Match 62.2%; Score 74; DB 2; Length 433;  
Best Local Similarity 57.1%; Pred. No. 0.0011;  
Matches 12; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 CNPSSGGRKVFELVGPSTYC 21  
DB 194 CMTDARGKLFNLVGPSTYC 214







```

C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 07-Aug-1998
C:Accession: G02913
R:Hara, T.
  submitted to GenBank, March 1996
A:Reference number: H01942
A:Accession: G02913
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-349 <HAR>
A:Cross-References: GB:D84105; NID:g1256700; PID:g1256701
C:Superfamily: complement factor H repeat homology
F:35-94/Domain: complement factor H repeat homology <FH01>
F:162-223/Domain: complement factor H repeat homology <FH03>
F:228-283/Domain: complement factor H repeat homology <FH02>

Query Match      42.4%  Score 50.5; DB 2; Length 349;
Best Local Similarity 47.6%  Pred. No. 3.1;
Matches 10; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Oy      1  CNPGSGGRKVFELVEGPSITC 21
      | | : | | | | | | | | |
Db      191 CDP-APGPPPSLVEGSTRVC 210

```

RESULT 15

B26359

decay-accelerating factor (GPI-anchored splice form) precursor - human

N:Alternate names: CD55; DAF splice form 2; decay-accelerating factor membrane-bound form

C:Species: Homo sapiens (man)

C:Date: 05-Oct-1988 #sequence\_revision 16-Aug-1996 #text\_change 24-Apr-1998

C:Accession: B26359; A27666; A39101; 152594; 152564

R:Caras, I.W.; Davitz, M.A.; Rhee, L.; Weddell, G.; Martin Jr., D.W.; Nussenzweig, V.

R:Nature 325, 545-549, 1987

A:Title: Cloning of decay-accelerating factor suggests novel use of splicing to generate

A:Reference number: A26359; MUID:87115845

A:Accession: B26359

A:Molecule type: mRNA

A:Residues: 1-381 <CAR>

A:Cross-references: GB:M30142; NID:g181464; PID:g181465

R:Medof, M.E.; Lublin, D.M.; Holers, V.M.; Ayers, D.J.; Getty, R.R.; Leykam, J.F.; Atkin

Proc. Natl. Acad. Sci. U.S.A. 84, 2007-2011, 1987

A:Title: Cloning and characterization of cDNAs encoding the complete sequence of decay-a

A:Reference number: A27666; MUID:87175602

A:Accession: A27666

A:Molecule type: mRNA

A:Residues: 6-79, 'T', '81-84, 'M', '86-381 <MED>

A:Cross-references: GB:M5799; NID:g181462; PID:g181463

R:Moran, P.; Raab, H.; Kohr, W.J.; Caras, I.W.

J. Biol. Chem. 266, 1250-1257, 1991

A:Title: Glycophospholipid membrane anchor attachment. Molecular analysis of the cleavag

A:Reference number: A39101; MUID:91093238

A:Accession: A39101

A:Molecule type: protein

A:Residues: 338-352 <MOR>

R:Lublin, D.M.; Mallinson, G.; Poole, J.; Reid, M.E.; Thompson, E.S.; Ferdman, B.R.; Tel

Blood 84, 1276-1282, 1994

A:Title: Molecular basis of reduced or absent expression of decay-accelerating factor in

A:Reference number: 152594

A:Accession: 152594

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 194-198, 'L', '200-209 <LUB>

A:Cross-references: GB:S72858; NID:g639599; PID:g639600

A:Experimental source: Individual KW, Cromer blood group phenotype Dr(a-)

A:Note: the single nucleotide difference in this allele, which changes Ser-199 to Leu, 1

on (see reference 152564), and thus reduced DAF expression

R:Reid, M.E.; Mallinson, G.; Slin, R.B.; Poole, J.; Pausch, V.; Merry, A.H.; Liew, Y.W.;

Blood 78, 3291-3297, 1991

A:Title: Biochemical studies on red blood cells from a patient with the Inab phenotype (

A:Reference number: 152564; MUID:92075980

A:Accession: 152564

A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 190-193, 'QLCPVE' <RE2>  
A:Cross-references: GB:S70688; NID:g240301; PID:g240302  
A:Experimental source: Individual KW, Cromer blood group phenotype Dr(a-) (described  
C:Comment: Cromer blood group system antigens reside on this protein.  
C:Genetics:  
A:Gene: GDB:DAF  
A:Cross-references: GDB:119088; OMIM:125240  
A:Map position: 1q32-1q32  
C:Function:  
A:Description: protects tissues from damage by regulating complement activation on cell  
C:Superfamily: decay-accelerating factor; complement factor H repeat homology  
C:Keywords: alternative splicing; complement inhibitor; glycoprotein; lipoprotein; ph  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:35-94/Product: decay-accelerating factor 2 #status predicted <MAT>  
F:36-94/Domain: complement factor H repeat homology <FH01>  
F:98-158/Domain: complement factor H repeat homology <FH02>  
F:163-220/Domain: complement factor H repeat homology <FH03>  
F:225-283/Domain: complement factor H repeat homology <FH04>  
F:354-381/Domain: carboxyl-terminal propeptide #status predicted <PRO>  
F:95/Blinding site: carboxylate (Asn) (covalent) #status predicted  
F:353/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Ser) (in mature f

```

Query Match Similarity      42.0%      Score 50: DB 1: Length 381.
Best Local Similarity      47.6%
Matches 10; Conservative 2; Mismatches 3; Indels 6; Gaps 1
QY      1 C N P G S G G R K V F L V G E P S I Y C 21
      11 |      | : : | | | | |
Db      253 C N R G - - - - - F T M I G H S I Y C 267

```

Search completed: July 18, 1999, 06:07:04  
Job time: 250 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 11:37:51 ; Search time 31.53 Seconds  
(without alignments)  
17.877 Million cell updates/sec

Title: US-09-142-043-4

Perfect score: 119  
Sequence: 1 CNPMSGGRKRVFELVGPSTIYC 21

Scoring table: BLOSUM62

Searched: 74019 seqs, 26840295 residues

Database: SwissProt\_36:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	119	100.0	2039	1 CRL_HUMAN	P17927 homo sapien
2	52	43.7	1033	1 CR2_HUMAN	P20023 homo sapien
3	50.5	42.4	377	1 MCP_HUMAN	P15529 homo sapien
4	50	42.0	381	1 DAF_HUMAN	P08174 homo sapien
5	49	41.2	1025	1 CR2_MOUSE	P19070 mus musculu
6	47	39.5	340	1 DAF_PONPY	P49457 pongo pygma
7	46	38.7	1786	1 LMB1_HUMAN	P07942 homo sapien
8	46	38.7	1786	1 LMB1_MOUSE	P02469 mus musculu
9	45.5	38.2	271	1 S24P_ANOGA	Q17004 anopheles g
10	45	37.8	469	1 C4B8_MOUSE	P08607 mus musculu
11	45	37.8	217	1 RPR_HELPR	P56188 heliococtae
12	44.5	37.4	476	1 HGDN_ACTIF	P11569 acidamococ
13	44	37.0	589	1 CAH_DUNSA	P54212 dunaliella
14	44	37.0	390	1 DAF1_MOUSE	O61475 mus musculu
15	44	37.0	407	1 DAF2_MOUSE	O61476 mus musculu
16	44	37.0	243	1 NOET_RHIN	P55473 rhizobium s
17	43.5	36.6	156	1 BOPD_HUMAN	P36639 homo sapien
18	43.5	36.6	321	1 SOLI_YEAST	P50578 saccharomyc
19	43	36.1	274	1 T2B1_HERAU	P25257 herpetosiph
20	43	36.1	274	1 T2B1_HERAU	P25260 herpetosiph
21	43	36.1	414	1 YBEC_BACSU	P40407 bacillus su
22	42	35.3	507	1 DAF_CAYPO	O60401 cavia porce
23	42	35.3	158	1 DYRE_STRAU	P10167 staphylococ
24	42	35.3	165	1 DYRC_ECOLI	O59408 escherichia
25	42	35.3	222	1 XYN2_TRIRE	P36217 trichoderma
26	42	35.3	190	1 XYN2_TRIRE	P48793 trichoderma
27	41.5	34.9	615	1 COX1_BACPC	P16262 bacillus ps
28	41.5	34.9	1765	1 POLN_PCVCP	P27407 feline gall
29	41.5	34.9	401	1 YQV1_CAEEL	O09551 caenorhabdi
30	41	34.5	548	1 ALGB_YEAST	P53354 saccharomyc
31	41	34.5	333	1 CBHR_XANFL	P25545 xanthobacte
32	41	34.5	168	1 DYR_BACSU	P11045 bacillus su
33	41	34.5	160	1 DYR_HAEIN	P43791 haemophilus
34	41	34.5	1799	1 LMB2_MOUSE	O61292 mus musculu
35	41	34.5	1801	1 LMB2_RAT	P13800 rattus norv
36	41	34.5	263	1 VCP_VACCV	P10998 vaccinia vi
37	41	34.5	225	1 XYN2_ASPNG	P55330 aspergillus
38	41	34.5	1522	1 YB6A_SCHPO	O09750 schizosacch
39	40.5	34.0	675	1 CA39_CHICK	P32017 gallus gall
40	40.5	34.0	360	1 FIYB_AROVI	P53574 azotobacter
41	40	33.6	454	1 AATK_LUPAN	P26563 lupinus ang
42	40	33.6	883	1 APCE_CYAPA	P48088 cyanophora
43	40	33.6	896	1 APCE_SYNY4	O02907 synecocyst

## ALIGNMENTS

RESULT	1	44	40	33.6	958	1	HIG_DROME
CRL_HUMAN							
ID	CRL_HUMAN						
AC	P17927;						
DT	01-NOV-1990 (REL. 16, CREATED)						
DT	01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)						
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)						
DE	COMPLEMENT RECEPTOR TYPE 1 PRECURSOR (C3B/C4B RECEPTOR) (CD35 ANTIGEN)						
GN	CRI OR C3BR.						
OS	HOMO SAPIENS (HUMAN).						
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;						
OC	EUTHERIA; PRIMATES.						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RX	MEDLINE: 89035992.						
RA	KLICKSTEIN L.B., BARTOW T.J., MILETIC V., RABSON L.D., SMITH J.A.,						
RA	FEARON D.T.;						
RL	J. EXP. MED. 168:1699-1717(1988).						
RN	[2]						
RP	SEQUENCE OF 503-2039 FROM N.A.						
RX	MEDLINE: 87168191.						
RA	KLICKSTEIN L.B., WONG W.W., SMITH J.A., WEIS J.H., WILSON J.G.,						
RA	FEARON D.T.;						
RL	J. EXP. MED. 165:1095-1112(1987).						
RN	[3]						
RP	SEQUENCE OF 761-783; 831-845 AND 1179-1195 FROM N.A.						
RX	MEDLINE: 86067975.						
RA	WONG W.W., KLICKSTEIN L.B., SMITH J.A., WEIS J.H., FEARON D.T.;						
RA	PROC. NATL. ACAD. SCI. U.S.A. 82:7711-7715(1985).						
CC	- FUNCTION: CRI, PRESENT ON ERYTHROCYTES, LEUKOCYTES, GLOMERULAR						
CC	PODOCYTES, AND SPLENIC FOLLICULAR DENDRITIC CELLS, MEDIATES THE						
CC	BINDING BY THESE CELLS OF PARTICLES AND IMMUNE COMPLEXES THAT HAVE						
CC	ACTIVATED COMPLEMENT.						
CC	- SUBUNIT: MONOMER.						
CC	- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.						
CC	- POLYPEPTIDISM: CRI IS RESPONSIBLE FOR THE KNOBS BLOOD GROUP SYSTEM.						
CC	- SEVEN SHORT CONSENSUS REPEATS (SCR) CONSTITUTE A LONG HOMOLOGOUS						
CC	REPEAT (LHR). THE N-TERMINAL TWO SCRS OF LHR-A CONTAINED A SITE						
CC	DETERMINING C4 SPECIFICITY, AND THE N-TERMINAL TWO SCRS OF LHR-B						
CC	- THIS IS THE SEQUENCE OF THE F ALLOTYPIC OF CRI.						
CC	- SIMILARITY: CONTAINS 30 SUSHI (SCR) REPEATS.						
DR	EMBL: M11617; G180996; -						
DR	EMBL: M11617; G180996; -						
DR	EMBL: Y00816; G30186; -						
DR	EMBL: X05309; G809019; -						
DR	PIR: A28507; A28507; -						
DR	PIR: A24748; A24748; -						
DR	PIR: B24748; B24748; -						
DR	PIR: C24748; C24748; -						
DR	PIR: S03843; S03843; -						
DR	HSSP: P08603; IHFI.						
DR	MIM: 120620; -						
KW	COMPLEMENT PATHWAY; GLYCOPROTEIN; TRANSMEMBRANE; REPEAT; SIGNAL;						
KW	RECEPTOR; SUSHI; BLOOD GROUP ANTIGEN.						
FT	SIGNAL	1	41				
FT	CHAIN	42	2039				
FT	DOMAIN	42	1971				
FT	DOMAIN	1972	1996				
FT	DOMAIN	1997	2039				
FT	MOD.RES	42	42				
FT	DOMAIN	42	488				
FT	REPEAT	42	100				
FT	SUSHI AL.						

```

FT REPEAT 103 162 SUSHI A2.
FT REPEAT 165 233 SUSHI A3.
FT REPEAT 237 294 SUSHI A4.
FT REPEAT 296 354 SUSHI A5.
FT REPEAT 357 417 SUSHI A6.
FT REPEAT 420 488 SUSHI A7.
FT DOMAIN 492 938 7 x SUSHI
FT REPEAT 492 938 REPEAT B.
FT REPEAT 550 550 SUSHI B1.
FT REPEAT 553 612 SUSHI B2.
FT REPEAT 615 683 SUSHI B3.
FT REPEAT 687 744 SUSHI B4.
FT REPEAT 746 804 SUSHI B5.
FT REPEAT 807 867 SUSHI B6.
FT REPEAT 938 938 SUSHI B7.
FT DOMAIN 942 1388 7 x SUSHI
FT REPEAT 942 1388 REPEAT C.
FT REPEAT 942 1000 SUSHI C1.
FT REPEAT 1003 1062 SUSHI C2.
FT REPEAT 1065 1133 SUSHI C3.
FT REPEAT 1137 1194 SUSHI C4.
FT REPEAT 1196 1254 SUSHI C5.
FT REPEAT 1257 1317 SUSHI C6.
FT REPEAT 1320 1388 SUSHI C7.
FT DOMAIN 1395 1846 7 x SUSHI
FT REPEAT 1395 1846 REPEAT D.
FT REPEAT 1395 1453 SUSHI D1.
FT REPEAT 1456 1515 SUSHI D2.
FT REPEAT 1518 1586 SUSHI D3.
FT REPEAT 1590 1647 SUSHI D4.
FT REPEAT 1649 1707 SUSHI D5.
FT REPEAT 1710 1770 SUSHI D6.
FT REPEAT 1773 1841 SUSHI D7.
FT DOMAIN 1847 1966 2 x SUSHI
FT REPEAT 1847 1966 SUSHI (SCR) REPEATS.
FT REPEAT 1847 1905 SUSHI 1.
FT REPEAT 1908 1966 SUSHI 2.
FT DISULFID 43 86 BY SIMILARITY.
FT DISULFID 73 99 BY SIMILARITY.
FT DISULFID 104 145 BY SIMILARITY.
FT DISULFID 131 161 BY SIMILARITY.
FT DISULFID 166 215 BY SIMILARITY.
FT DISULFID 195 232 BY SIMILARITY.
FT DISULFID 238 280 BY SIMILARITY.
FT DISULFID 266 293 BY SIMILARITY.
FT DISULFID 297 340 BY SIMILARITY.
FT DISULFID 326 353 BY SIMILARITY.
FT DISULFID 358 400 BY SIMILARITY.
FT DISULFID 386 416 BY SIMILARITY.
FT DISULFID 421 470 BY SIMILARITY.
FT DISULFID 450 487 BY SIMILARITY.
FT DISULFID 493 536 BY SIMILARITY.
FT DISULFID 523 549 BY SIMILARITY.
FT DISULFID 554 595 BY SIMILARITY.
FT DISULFID 581 611 BY SIMILARITY.
FT DISULFID 616 665 BY SIMILARITY.
FT DISULFID 645 682 BY SIMILARITY.
FT DISULFID 688 730 BY SIMILARITY.
FT DISULFID 716 743 BY SIMILARITY.
FT DISULFID 747 790 BY SIMILARITY.
FT DISULFID 776 803 BY SIMILARITY.
FT DISULFID 808 850 BY SIMILARITY.
FT DISULFID 836 866 BY SIMILARITY.
FT DISULFID 871 920 BY SIMILARITY.
FT DISULFID 900 937 BY SIMILARITY.
FT DISULFID 943 986 BY SIMILARITY.
FT DISULFID 973 999 BY SIMILARITY.
FT DISULFID 1004 1045 BY SIMILARITY.
FT DISULFID 1031 1061 BY SIMILARITY.
FT DISULFID 1066 1115 BY SIMILARITY.
FT DISULFID 1095 1132 BY SIMILARITY.
FT DISULFID 1138 1180 BY SIMILARITY.
FT DISULFID 1166 1193 BY SIMILARITY.
FT DISULFID 1197 1240 BY SIMILARITY.

```

```

FT DISULFID 1226 1253 BY SIMILARITY.
FT DISULFID 1258 1300 BY SIMILARITY.
FT DISULFID 1286 1316 BY SIMILARITY.
FT DISULFID 1321 1370 BY SIMILARITY.
FT DISULFID 1350 1387 BY SIMILARITY.
FT DISULFID 1396 1439 BY SIMILARITY.
FT DISULFID 1426 1452 BY SIMILARITY.
FT DISULFID 1457 1498 BY SIMILARITY.
FT DISULFID 1484 1514 BY SIMILARITY.
FT DISULFID 1519 1568 BY SIMILARITY.
FT DISULFID 1548 1585 BY SIMILARITY.
FT DISULFID 1591 1633 BY SIMILARITY.
FT DISULFID 1619 1646 BY SIMILARITY.
FT DISULFID 1650 1693 BY SIMILARITY.
FT DISULFID 1679 1706 BY SIMILARITY.
FT DISULFID 1711 1753 BY SIMILARITY.
FT DISULFID 1739 1769 BY SIMILARITY.
FT DISULFID 1774 1823 BY SIMILARITY.
FT DISULFID 1803 1840 BY SIMILARITY.
FT DISULFID 1848 1891 BY SIMILARITY.
FT DISULFID 1877 1904 BY SIMILARITY.
FT DISULFID 1909 1952 BY SIMILARITY.
FT DISULFID 1938 1965 BY SIMILARITY.
FT CARBOHYD 36 56 POTENTIAL.
FT CARBOHYD 252 252 POTENTIAL.
FT CARBOHYD 410 410 POTENTIAL.
FT CARBOHYD 447 447 POTENTIAL.
FT CARBOHYD 509 509 POTENTIAL.
FT CARBOHYD 578 578 POTENTIAL.
FT CARBOHYD 702 702 POTENTIAL.
FT CARBOHYD 860 860 POTENTIAL.
FT CARBOHYD 897 897 POTENTIAL.
FT CARBOHYD 959 959 POTENTIAL.
FT CARBOHYD 1028 1028 POTENTIAL.
FT CARBOHYD 1152 1152 POTENTIAL.
FT CARBOHYD 1310 1310 POTENTIAL.
FT CARBOHYD 1481 1481 POTENTIAL.
FT CARBOHYD 1504 1504 POTENTIAL.
FT CARBOHYD 1534 1534 POTENTIAL.
FT CARBOHYD 1540 1540 POTENTIAL.
FT CARBOHYD 1605 1605 POTENTIAL.
FT CARBOHYD 1763 1763 POTENTIAL.
FT CARBOHYD 1908 1908 POTENTIAL.
SQ SEQUENCE 2039 AA; 223589 MW; 666F9033 CRC32;

Query Match 100.0%; Score 119; DB 1; Length 2039;
Best Local Similarity 100.0%; Pred. No. 2.9e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVELVGEPSIYC 21
| | | | | | | | | | | | | | | | | | | | |
Db 195 CNPSSGGRKVELVGEPSIYC 215

RESULT 2
CR2_HUMAN STANDARD; PRT; 1033 AA.
ID CR2_HUMAN
AC P20023;
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE COMPLEMENT RECEPTOR TYPE 2 PRECURSOR (CR2) (COMPLEMENT C3D RECEPTOR)
DE (PESSTEIN-BARR VIRUS RECEPTOR) (CD21 ANTIGEN).
GN CR2 OR C3DR.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89123277.
RA FUJISAKU A., HARLEY J.B., FRANK M.B., GRUNER B.A., FRAZIER B.,
RA HOLERS V.M.;

```

RL J. BIOL. CHEM. 264:2118-2125(1989).  
 RN [2]  
 RP SEQUENCE OF 226-223; 256-267; 332-341; 667-677 AND 898-908.  
 RX MEDLINE; 86287311.  
 RA WEIS J.J., FEARON D.T., KLUCKSTEIN L.B., WONG W.W., RICHARDS S.A.,  
 DE BRUN KOPS A., SMITH J.A., WEIS J.H.;  
 CC PROC. NATL. ACADE. SCI. U.S.A. 83:5639-5643(1986).  
 CC - FUNCTION: RECEPTOR FOR COMPLEMENT C3D AND FOR THE EPSTEIN-BARR  
 CC VIRUS ON HUMAN B-CELLS. PARTICIPATES IN B LYMPHOCYTES ACTIVATION.  
 CC - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC - TISSUE SPECIFICITY: MATURE B LYMPHOCYTES AND FOLLICULAR  
 CC DENDRITIC CELLS OF THE SPLEEN.  
 CC - SIMILARITY: TO MOUSE CR2. CD21 IS A MEMBER OF THE REGULATORS OF  
 CC COMPLEMENT ACTIVATION (RCA) FAMILY.  
 CC - SIMILARITY: CONTAINS 15 SUSHI (SCR) REPEATS.  
 CC - DATABASE: NAME-PROW; NOTE-CD guide CD21 entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd21.htm".  
 DR EMBL; M26004; G181940; -.  
 DR EMBL; M26016; G181938; -.  
 DR EMBL; M24007; G181938; JOINED.  
 DR EMBL; M24008; G181938; JOINED.  
 DR EMBL; M24009; G181938; JOINED.  
 DR EMBL; M24010; G181938; JOINED.  
 DR EMBL; M24011; G181938; JOINED.  
 DR EMBL; M26009; G181938; JOINED.  
 DR EMBL; M26010; G181938; JOINED.  
 DR EMBL; M26011; G181938; JOINED.  
 DR EMBL; M26012; G181938; JOINED.  
 DR EMBL; M26013; G181938; JOINED.  
 DR EMBL; M26014; G181938; JOINED.  
 DR EMBL; M26015; G181938; JOINED.  
 DR PIR; A32036; A32036.  
 DR PIR; A24319; A24319.  
 DR PIR; B24319; B24319.  
 DR PIR; C24319; C24319.  
 DR PIR; D24319; D24319.  
 DR PIR; E24319; E24319.  
 DR PIR; F24319; F24319.  
 DR HSSP; P08603; 1HFI.  
 DR MIM; L10650; -.  
 KW COMPLEMENT PATHWAY; GLYCOPROTEIN; TRANSMEMBRANE; REPEAT; SIGNAL;  
 KM RECEPTOR; SUSHI.  
 FT SIGNAL 1 20  
 FT CHAIN 21 1033 COMPLEMENT RECEPTOR TYPE 2.  
 FT DOMAIN 21 971 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 972 999 POTENTIAL.  
 FT DOMAIN 1000 1033 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 22 971 15 X SUSHI (SCR) REPEATS.  
 FT DOMAIN 22 83 SUSHI 1.  
 FT REPEAT 22 83 SUSHI 2.  
 FT REPEAT 90 147 SUSHI 3.  
 FT REPEAT 153 211 SUSHI 4.  
 FT REPEAT 214 272 SUSHI 5.  
 FT REPEAT 275 343 SUSHI 6.  
 FT REPEAT 350 407 SUSHI 7.  
 FT REPEAT 409 467 SUSHI 8.  
 FT REPEAT 470 523 SUSHI 9.  
 FT REPEAT 526 594 SUSHI 10.  
 FT REPEAT 601 658 SUSHI 11.  
 FT REPEAT 661 715 SUSHI 12.  
 FT REPEAT 718 780 SUSHI 13.  
 FT REPEAT 787 844 SUSHI 14.  
 FT REPEAT 850 908 SUSHI 15.  
 FT REPEAT 911 969 SUSHI 15.  
 FT DISULFID 23 65 BY SIMILARITY.  
 FT DISULFID 51 82 BY SIMILARITY.  
 FT DISULFID 91 132 BY SIMILARITY.  
 FT DISULFID 118 146 BY SIMILARITY.  
 FT DISULFID 154 197 BY SIMILARITY.  
 FT DISULFID 183 210 BY SIMILARITY.  
 FT DISULFID 215 256 BY SIMILARITY.  
 FT DISULFID 242 271 BY SIMILARITY.  
 FT DISULFID 276 325 BY SIMILARITY.  
 FT DISULFID 305 342 BY SIMILARITY.

FT DISULFID 351 393 BY SIMILARITY.  
 FT DISULFID 379 406 BY SIMILARITY.  
 FT DISULFID 410 453 BY SIMILARITY.  
 FT DISULFID 439 466 BY SIMILARITY.  
 FT DISULFID 471 509 BY SIMILARITY.  
 FT DISULFID 495 522 BY SIMILARITY.  
 FT DISULFID 527 576 BY SIMILARITY.  
 FT DISULFID 556 593 BY SIMILARITY.  
 FT DISULFID 602 644 BY SIMILARITY.  
 FT DISULFID 630 657 BY SIMILARITY.  
 FT DISULFID 662 699 BY SIMILARITY.  
 FT DISULFID 685 714 BY SIMILARITY.  
 FT DISULFID 719 762 BY SIMILARITY.  
 FT DISULFID 748 779 BY SIMILARITY.  
 FT DISULFID 788 830 BY SIMILARITY.  
 FT DISULFID 816 843 BY SIMILARITY.  
 FT DISULFID 851 894 BY SIMILARITY.  
 FT DISULFID 880 907 BY SIMILARITY.  
 FT DISULFID 912 955 BY SIMILARITY.  
 FT DISULFID 941 968 BY SIMILARITY.  
 FT CARBOHYD 121 121 POTENTIAL.  
 FT CARBOHYD 127 127 POTENTIAL.  
 FT CARBOHYD 294 294 POTENTIAL.  
 FT CARBOHYD 372 372 POTENTIAL.  
 FT CARBOHYD 492 492 POTENTIAL.  
 FT CARBOHYD 623 623 POTENTIAL.  
 FT CARBOHYD 682 682 POTENTIAL.  
 FT CARBOHYD 800 800 POTENTIAL.  
 FT CARBOHYD 823 823 POTENTIAL.  
 FT CARBOHYD 861 861 POTENTIAL.  
 FT CARBOHYD 911 911 POTENTIAL.  
 FT CONFLICT 667 667 O -> D (IN REF. 2).  
 FT CONFLICT 902 902 O -> G (IN REF. 2).  
 FT CONFLICT 906 906 H -> L (IN REF. 2).  
 SQ SEQUENCE 1033 AA; 112973 MW; 7032F558 CRC32;

Query Match 43.7%; Score 52; DB 1; Length 1033;  
 Best Local Similarity 47.6%; Pred. No. 2.7;  
 Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Oy 1 CNPSSGGRKRVFELVGPSTYC 21  
 Db 556 CNPGRGVEFSLIGESTIRC 576

RESULT 3  
 MCP\_HUMAN STANDARD; PRT; 377 AA.  
 AC P15529;  
 DT 01-APR-1990 (REL. 14, CREATED)  
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
 DE MEMBRANE COFACTOR PROTEIN PRECURSOR (CD46 ANTIGEN) (TROPHOBLAST  
 DE LEUCOCYTE COMMON ANTIGEN) (TLX).  
 GN MCP.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 35-58.  
 RX LUBLIN D.M., LISZEWski M.K., POST T.W., ARCE M.A., LE BEAU M.M.,  
 RA REBERTSCH M.B., LEMONS R.S., SETA T., ATKINSON J.P.;  
 RL J. EXP. MED. 168:181-194(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE-TESTIS.  
 RX MEDLINE: 93119658.  
 RA CERVONI F., FENICHEL P., AKHOUNDI C., HSI B.L., ROSSI B.;  
 RL MOL. REPROD. DEV. 34:107-113(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.

RX MEDLINE; 91267562.  
 RA PORCELL D.F., RUSSELL S.M., DEACON N.J., BROWN M.A., HOOKER D.J.,  
 RA MCKENZIE I.F.;  
 RN IMMUNOGENETICS 33:335-344(1991).  
 RP [4]  
 RP SEQUENCE OF 1-34 FROM N.A.  
 RX MEDLINE; 94014356.  
 RA CUI W., HOUCADE D., POST T., GREENLUND A.C., ATKINSON J.P.,  
 RA KUMAR V.;  
 RL J. IMMUNOL. 151:4137-4146(1993).  
 RN [5]  
 RP ALTERNATIVE SPLICING.  
 RX MEDLINE; 92289809.  
 RA RUSSELL S.M., SPARROW R.L., MCKENZIE I.F.C., PORCELL D.F.J.;  
 RL EUR. J. IMMUNOL. 22:1513-1518(1992).  
 CC -1- FUNCTION: MCP IS AN IMPORTANT MEMBRANE PROTEIN FOR PROTECTING HOST  
 CC CELLS FROM DAMAGE BY COMPLEMENT. MCP POSSESSES COFACTOR ACTIVITY  
 CC FOR INACTIVATION OF C3B AND C4B BY SERUM FACTOR I. THIS WOULD  
 CC ALLOW MCP TO REGULATE THE COMPLEMENT SYSTEM AT THE LEVEL OF THE C3  
 CC CONVERTASES. MAY BE INVOLVED IN THE FUSION OF THE SPERMATOZOA WITH  
 CC THE OOCYTE DURING FERTILIZATION. MAY ACT BY PREVENTING COMPLEMENT  
 CC ACTIVATION AT THE FETO-MATERNAL INTERFACE ON THE  
 CC SYNCYTOTROPHOBLAST LAYER OF PLACENTA.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED BY ALMOST ALL CELLS EXCEPT  
 CC ERYTHROCYTES AND SOME BONE MARROW CELLS.  
 CC -1- PFM: EXTENSIVELY O-GLYCOSYLATED IN THE SER/THR-RICH DOMAIN  
 CC (PROBABLE).  
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 14 FORMS ARE PRODUCED BY  
 CC ALTERNATIVE SPLICING. THE FORM SHOWN IS THE TRANSCRIPT A. THE  
 CC GAMMA ISOFORM (TRANSCRIPTS A AND B) IS PREFERENTIALLY EXPRESSED IN  
 CC BAV-B CELLS AND LEUKEMIC CELLS. THE 66 KD ALPHA ISOFORM AND THE 56  
 CC KD BETA ISOFORM FOUND IN ALL TISSUES EXCEPT SPERM CORRESPOND TO  
 CC THE TRANSCRIPTS C AND D, AND THE TRANSCRIPTS E AND F RESPECTIVELY.  
 CC THE EXON 9 IS SPECIFICALLY DELETED IN SOME PLACENTAL. THE EPSILON  
 CC ISOFORM CORRESPONDS TO THE TRANSCRIPTS I AND J. SPERMATOZOA DELTA  
 CC ISOFORM (TRANSCRIPT N) DELETE BOTH EXONS 12 AND 13. ALL TISSUES  
 CC DIFFERENTIALLY SPLICED EXON 13.  
 CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.  
 CC -1- DATABASE: NAME-PROW; NOTE-CD guide cd46 entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd46.htm".  
 CC EMBL; Y00651; G34505; -;  
 DR EMBL; S51940; G262938; -;  
 DR EMBL; M58050; G180137; -;  
 DR EMBL; A18585; G512457; -;  
 DR EMBL; S65879; E91388; -;  
 DR PIR; S01896; S01896.  
 DR MIM; 120920; -;  
 KW COMPLEMENT PATHWAY; GLYCOPROTEIN; TRANSMEMBRANE; REPEAT; SIGNAL;  
 KW SUSHI; ALTERNATIVE SPLICING.  
 FT SIGNAL 1 34  
 FT CHAIN 35 377 MEMBRANE COFACTOR PROTEIN.  
 FT DOMAIN 35 328 EXTRACELLULAR (POTENTIAL).  
 FT TRANSERM 329 351 POTENTIAL.  
 FT DOMAIN 352 377 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 35 284 4 X SUSHI (SCR) REPEATS.  
 FT REPEAT 35 95 SUSHI 1.  
 FT REPEAT 98 158 SUSHI 2.  
 FT REPEAT 161 224 SUSHI 3.  
 FT REPEAT 227 284 SUSHI 4.  
 FT DOMAIN 287 311 SER/THR-RICH.  
 FT DISULFID 35 80 BY SIMILARITY.  
 FT DISULFID 64 94 BY SIMILARITY.  
 FT DISULFID 99 141 BY SIMILARITY.  
 FT DISULFID 127 157 BY SIMILARITY.  
 FT DISULFID 162 210 BY SIMILARITY.  
 FT DISULFID 191 223 BY SIMILARITY.  
 FT DISULFID 228 270 BY SIMILARITY.  
 FT DISULFID 256 283 BY SIMILARITY.  
 FT CARBOHYD 83 83 POTENTIAL.  
 FT CARBOHYD 114 114 POTENTIAL.  
 FT CARBOHYD 273 273 POTENTIAL.  
 FT VARSPLIC 286 300 MISSING (IN A 2ND FORM).

FT VARSPLIC 340 352 MISSING (IN TRANSCRIPT M).  
 FT VARSPLIC 353 361 YLORRRAK -> DIRKGRKQKWEIMPLRLNPLQ  
 FT VARSPLIC 362 377 OSREAE (IN TRANSCRIPT M).  
 FT VARSPLIC 362 377 TYLDETHREVFTS -> KADGAEVATYQKSTPAEQ  
 FT VARSPLIC 362 377 RG (IN TRANSCRIPTS B, D, F, H, J, AND L).  
 FT VARSPLIC 340 361 MISSING (IN TRANSCRIPT M AND N).  
 FT VARSPLIC 339 339 MISSING (IN TRANSCRIPT N).  
 FT VARSPLIC 339 339 I -> IKGOMVELMPLRLNPLQOSREAE (IN  
 TRANSCRIPT N).  
 SQ SEQUENCE 377 AA; 42247 MM; 7FE9E38E CRC32;  
 Query Match 42.4%; Score 50.5; DB 1; Length 377;  
 Best Local Similarity 47.6%; Pred. No. 1.7;  
 Matches 10; Conservative 4; Mismatches 6; Indels 1; Gaps 1;  
 QY 1 CNGSCGRKVFELVGEPTIC 21  
 Db 191 CDP-APGDPFSLIGESTIVC 210  
 DAF HUMAN  
 ID DAF\_HUMAN STANDARD; PRT; 381 AA.  
 AC P08174; P09679;  
 DT 01-AUG-1988 (REL. 08, CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE COMPLEMENT DECAY-ACCELERATING FACTOR PRECURSOR (CD55).  
 GN DAF OR CD55.  
 OS HOMO SAPIENS (HUMAN).  
 OC ECHINOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 87115845.  
 RA CARAS I.W., DAVITZ M.A., RHEE L., WEDDELL G., MARTIN D.W. JR.,  
 RA NISSENZWEIG V.;  
 RL NATURE 325:545-549(1987).  
 RN [2]  
 RP SEQUENCE OF 6-381 FROM N.A. (DAF-2 FORM).  
 RX MEDLINE; 87175602.  
 RA MEDOF M.E., LUBLIN D.M., HOLERS V.M., AYERS D.J., GETTY R.R.,  
 RA LEYKAM J.F., ATKINSON J.P., TYKOCINSKI M.L.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 84:2007-2011(1987).  
 RN [3]  
 RP SEQUENCE OF 1-100 FROM N.A.  
 RX MEDLINE; 91271256.  
 RA EWULONU U.K., RAVI L., MEDOF M.E.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 88:4675-4679(1991).  
 RN [4]  
 RP SEQUENCE OF 35-46.  
 RC TISSUE-URINE.  
 RX MEDLINE; 91291869.  
 RA NAKANO Y., SUGITA Y., ISHIKAWA Y., CHOI N.-H., TOBE T., TOMITA M.;  
 RL BIOCHIM. BIOPHYS. ACTA 1074:326-330(1991).  
 RN [5]  
 RP GPI-ANCHOR.  
 RX MEDLINE; 91093238.  
 RA MORAN P., RAAB H., KOHR W.J., CARAS I.W.;  
 RL J. BIOL. CHEM. 266:1250-1257(1991).  
 RN [6]  
 RP DISULFIDE BONDS IN SUSHI DOMAINS.  
 RX MEDLINE; 92305034.  
 RA NAKANO Y., SUMIDA K., KIKUTA N., MURA N.-H., TOBE T., TOMITA M.;  
 RL BIOCHIM. BIOPHYS. ACTA 1116:235-240(1992).  
 RN [7]  
 RP FUNCTION AS A ECHOVIRUS RECEPTOR.  
 RX MEDLINE; 95045399.  
 RA WARD T., PIRKIN P.A., CLARKSON N.A., STONE D.M., MINOR P.D.,  
 RA ALMOND J.W.;  
 RL EMBO J. 13:5070-5074(1994).  
 RN [8]

RP	VARIANT BLOOD GROUP DR(A-).
RX	MEDLINE: 94325573.
RA	LUBLIN D.M., MALLINSON G., POOLE J., REID M.E., THOMPSON E.S.,
RA	FEDMAN B.R., TELEN M.J., ANSTEE D.J., TANNER M.J.A.;
RL	BLOOD 84;1:1276-1282(1994).
CC	-1- FUNCTION: THIS PROTEIN RECOGNIZES C4B AND C3B FRAGMENTS THAT
CC	CONDENSE WITH CELL-SURFACE HYDROXYL OR AMINO GROUPS WHEN NASCENT
CC	C4B AND C3B ARE LOCALLY GENERATED DURING C4 AND C3 ACTIVATION.
CC	INTERFERES WITH THEIR ABILITY TO CATALYZE THE CONVERSION OF C2 AND
CC	FACTOR B TO ENZYMATICALLY ACTIVE C2A AND BB AND THEREBY PREVENTS
CC	THE FORMATION OF C4B2A AND C3BBB, THE AMPLIFICATION CONVERGENCE OF
CC	THE COMPLEMENT CASCADE.
CC	-1- FUNCTION: ALSO ACTS AS THE RECEPTOR FOR ECCHOVIRUS 7 AND RELATED
CC	VIRUSES (ECCHOVIRUSES 13, 21, 29 AND 33).
CC	-1- SUBUNIT: MONOMER (MAJOR FORM) AND NON-DISULFIDE-LINKED, COVALENT
CC	HOMODIMER (MINOR FORM).
CC	-1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC	-1- TISSUE SPECIFICITY: EXPRESSED ON THE PLASMA MEMBRANES OF ALL CELL
CC	TYPES THAT ARE IN INTIMATE CONTACT WITH PLASMA COMPLEMENT
CC	PROTEINS. IT IS ALSO FOUND ON THE SURFACES OF EPITHELIAL CELLS
CC	LINING EXTRACELLULAR COMPARTMENTS, AND VARIANTS OF THE MOLECULE
CC	ARE PRESENT IN BODY FLUIDS AND IN EXTRACELLULAR MATRIX.
CC	-1- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR
CC	FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE
CC	ACTIVE SITE ON SCR3 (BY SIMILARITY).
CC	-1- PTM: THE SER/THR-RICH DOMAIN IS HEAVILY O-GLYCOSYLATED.
CC	-1- POLYMORPHISM: DAF IS RESPONSIBLE FOR THE CHROMER BLOOD GROUP
CC	SYSTEM. IT CONSISTS OF AT LEAST SEVEN HIGH-INCIDENCE (CR(A),
CC	TC(B), DR(A), ES(A), WE(S)(B), UMC, AND IFC) AND LOW-INCIDENCE
CC	(TG(C)(B), TC(C), AND WE(S)(A)) ANTIGENS THAT RESIDE ON DAF. IN THE
CC	CHROMER PHENOTYPES DR(A-) AND INAB THERE IS REDUCED OR ABSENT
CC	EXPRESSION OF DAF, RESPECTIVELY. IN THE CASE OF THE DR(A-)
CC	PHENOTYPE, A SINGLE NUCLEOTIDE SUBSTITUTION WITHIN EXON 5 ACCOUNTS
CC	FOR TWO CHANGES: A SIMPLE AMINO ACID SUBSTITUTION THAT IS THE
CC	BASIS OF THE ANTIGENIC VARIATION, AND AN ALTERNATIVE SPLICING
CC	EVEN THAT UNDERLIES THE DECREASED EXPRESSION OF DAF IN THIS
CC	PHENOTYPE.
CC	-1- ALTERNATIVE PRODUCTS: TWO FORMS OF DAF (DAF-2, SHOWN HERE, AND
CC	DAF-1) ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.
CC	-1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.
CC	-1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION
CC	(RCA) FAMILY.
CC	-1- DATABASE: NAME=PROW; NOTE=CD guide CD5 entry;
CC	WWW=http://www.ncbi.nlm.nih.gov/prov/cd/cd5.htm*.
DR	EMBL: M31516; GI81468; -.
DR	EMBL: M30142; GI81465; -.
DR	EMBL: M15799; GI81463; -.
DR	EMBL: M64653; GI81476; -.
DR	EMBL: M64356; GI81476; JOINED.
DR	EMBL: S72858; G639600; -.
DR	PIR: B26359; B26359.
DR	PIR: A26359; A26359.
DR	PIR: S16187; S16187.
DR	PIR: A39101; A39101.
DR	PIR: S23138; S23138.
DR	HSSP: P08603; IHFI.
DR	KIM: L25240; -.
KM	ALTERNATE PATHWAY: PLASMA: GLYCOPROTEIN. MEMBRANE: REPEAT:
KM	ALTERNATIVE SPLICING; GPI-ANCHOR; SIGNAL; SUSHI; POLYMORPHISM;
KM	BLOOD GROUP ANTIGEN.
FT	SIGNAL 1 34
FT	CHAIN 35 353
FT	PROPEP 354 381
FT	DOMAIN 35 284
FT	REPEAT 35 95
FT	REPEAT 97 159
FT	REPEAT 162 221
FT	REPEAT 224 284
FT	DOMAIN 287 356
FT	DISEUFLD 36 81
FT	DISEUFLD 65 94
FT	DISEUFLD 98 145

FT	DISULFID	129	158	
FT	DISULFID	163	204	
FT	DISULFID	190	220	
FT	DISULFID	225	267	
FT	DISULFID	253	283	
FT	CARBOHYD	95	95	
FT	LIPID	353	353	POTENTIAL.
FT	VARIANT	52	52	GPI-ANCHOR.
FT	VARIANT	52	52	R -> L (IN TC(B) ANTIGEN).
FT	VARIANT	82	82	R -> P (IN TC(C) ANTIGEN).
FT	VARIANT	199	199	L -> R (IN WES(A) ANTIGEN).
FT	VARIANT	227	227	S -> L (IN DR(A-) ANTIGEN).
FT	VARIANT	362	381	A -> P (IN CR(A-) ANTIGEN).
FT	VARIANT	362	381	HTCTLTGSLTLYTMGLT -> SRPTQAGMKCRDSSL
FT	CONFLICT	80	80	OSRPFGRSRSHSLPSWYRAVHFVDRFAMWASNNGLA
FT	CONFLICT	85	85	DLAKEELRRKTYQVRLFLVS (IN DAF-1).
FT	CONFLICT	85	85	T -> I (IN REF. 1).
FT	CONFLICT	85	85	S -> M (IN REF. 2).
SO	SEQUENCE	381 AA;	41388 MW;	114271A8 CRC32;

Query Match Best Local Similarity 42.0%; Score 50; DB 1; Length 381;  
Matches 10; Conservative 47.6%; Pred. No. 2;  
Matches 2; Mismatches 3; Indels 6; Gaps 1;

Qy 1 CNPSSGGRKVFELVGEPSIYC 21  
Db 253 CNKG-----FTMIGHSIYC 267

RESULT 5  
CR2\_MOUSE  
ID CR2\_MOUSE STANDARD: PRT: 1025 AA.  
AC P19070;  
DT 01-NOV-1990 (REL. 16, CREATED)  
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)  
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
DE COMPLEMENT RECEPTOR TYPE 2 PRECURSOR (CR2) (COMPLEMENT C3D RECEPTOR).  
GN CR2.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; ROSENTIA.  
RA [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN-BALB/C;  
RX MEDLINE: 90229735.  
RA FINGEROTH J.D.;  
RL J. IMMUNOL. 144:3458-3467(1990).  
RN [2]  
RN SEQUENCE OF 12-1025 FROM N.A.  
RX MEDLINE: 91010789.  
RA MOLINA H., KINOSHITA T., INOUE K., CAREL J.C., HOLERS V.M.;  
RL J. IMMUNOL. 145:2974-2983(1990).  
RN [3]  
RN SEQUENCE OF 343-401 AND 991-1025 FROM N.A.  
RX MEDLINE: 89098890.  
RA FINGEROTH J.D., BENEDICT M.A., LEVY D.N., STROMINGER J.L.;  
RL PROC. NATL. ACAD. SCI. U.S.A. 86:242-246(1989).  
RN [4]  
RN SEQUENCE OF 289-1025 FROM N.A.  
RX MEDLINE: 89381350.  
RA KUTZ C.B., PAUL M.S., AEGERTER M., WEIS J.J., WEIS J.H.;  
RL J. IMMUNOL. 143:2058-2067(1989).  
CC -1- FUNCTION: RECEPTOR FOR COMPLEMENT C3D. PARTICIPATES IN B LYMPHOCYTES ACTIVATION.  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -1- TISSUE SPECIFICITY: B LYMPHOCYTES.  
CC -1- SIMILARITY: TO HUMAN C3D/EBV RECEPTOR (CD21).  
CC -1- SIMILARITY: CONTAINS 15 SUSHI (SCR) REPEATS.  
DR EMBL: M81083; G192697; -  
DR EMBL: M35684; G192688; -  
DR EMBL: M61132; G192693; -  
DR EMBL: M35685; G192695; ALT\_SEQ.  
EMBL: M29281; G387131; -





FT CARBOHYD 107 107 POTENTIAL.  
LIPID 312 312 GPI-ANCHOR (BY SIMILARITY).  
SQ SEQUENCE 340 AA; 37180 MW; 5CAC18F8 CRC32;  
Query Match 39.58; Score 47; DB 1; Length 340;  
Best Local Similarity 42.98; Pred. No. 5.1;  
Matches 9; Conservative 3; Mismatches 3; Indels 6; Gaps 1;  
1 CNPMSGGRKVEELVGEPSIC 21  
| | | | |  
| | | | |  
Db 212 CNKG-----YTMGEHSITC 226  
RESULT 7  
LMB1\_HUMAN  
ID LMB1\_HUMAN STANDARD; PRT; 1786 AA.  
AC P07942;  
DT 01-AUG-1988 (REL. 08, CREATED)  
DT 01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE LAMININ BETA-1 CHAIN PRECURSOR (LAMININ B1 CHAIN).  
GN LAMB1  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
[1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 90368768.  
RA VOULTEENHO R., CHOW L.T., TRYGGVASON K.;  
RL J. BIOL. CHEM. 265:15611-15616(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 87280097.  
RA PIKKARAINEN T., EDDY R., FUKUSHIMA Y., BYERS M., SHOWS T.,  
RA PIHLAJANIEMI T., SARASTE M., TRYGGVASON K.;  
RL J. BIOL. CHEM. 262:10454-10462(1987).  
RN [3]  
RP SEQUENCE OF 1276-1709 FROM N.A.  
RX MEDLINE: 88021029.  
RA JAYE M., MODI W.S., RICCA G.A., MUDD R., CHIU I.M., O'BRIEN S.J.,  
RA DROHAN W.N.;  
RL AM. J. HUM. GENET. 41:605-615(1987).  
CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ  
IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF  
CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING  
WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.  
CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE  
DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND  
TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE  
COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.  
CC -1- SUBUNIT: THE BETA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ),  
LAMININ-2 (MEROSIN), AND LAMININ-6 (K-LAMININ).  
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR  
COMPONENT).  
CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT  
WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.  
CC -1- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.  
CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).  
CC -1- SIMILARITY: CONTAINS 12.5 LAMININ EGF-LIKE DOMAINS.  
DR EMBL: M61951; G186913; -  
DR EMBL: M58147; G186913; JOINED.  
DR EMBL: M61917; G186913; JOINED.  
DR EMBL: M61918; G186913; JOINED.  
DR EMBL: M61921; G186913; JOINED.  
DR EMBL: M61922; G186913; JOINED.  
DR EMBL: M61923; G186913; JOINED.  
DR EMBL: M61924; G186913; JOINED.  
DR EMBL: M61925; G186913; JOINED.  
DR EMBL: M61926; G186913; JOINED.  
DR EMBL: M61927; G186913; JOINED.

DR EMBL: M61928; G186913; JOINED.  
DR EMBL: M61929; G186913; JOINED.  
DR EMBL: M61930; G186913; JOINED.  
DR EMBL: M61931; G186913; JOINED.  
DR EMBL: M61932; G186913; JOINED.  
DR EMBL: M61933; G186913; JOINED.  
DR EMBL: M61934; G186913; JOINED.  
DR EMBL: M61935; G186913; JOINED.  
DR EMBL: M61936; G186913; JOINED.  
DR EMBL: M61937; G186913; JOINED.  
DR EMBL: M61938; G186913; JOINED.  
DR EMBL: M61939; G186913; JOINED.  
DR EMBL: M61940; G186913; JOINED.  
DR EMBL: M61941; G186913; JOINED.  
DR EMBL: M61942; G186913; JOINED.  
DR EMBL: M61943; G186913; JOINED.  
DR EMBL: M61944; G186913; JOINED.  
DR EMBL: M61945; G186913; JOINED.  
DR EMBL: M61946; G186913; JOINED.  
DR EMBL: M61947; G186913; JOINED.  
DR EMBL: M61948; G186913; JOINED.  
DR EMBL: M61949; G186913; JOINED.  
DR EMBL: M61950; G186913; JOINED.  
DR EMBL: M55370; G186876; -  
DR EMBL: M55378; G186876; JOINED.  
DR EMBL: M55365; G186876; JOINED.  
DR EMBL: M55371; G186876; JOINED.  
DR EMBL: M55372; G186876; JOINED.  
DR EMBL: M55373; G186876; JOINED.  
DR EMBL: M55374; G186876; JOINED.  
DR EMBL: M55375; G186876; JOINED.  
DR EMBL: M55376; G186876; JOINED.  
DR EMBL: M55344; G186876; JOINED.  
DR EMBL: M55345; G186876; JOINED.  
DR EMBL: M55346; G186876; JOINED.  
DR EMBL: M55347; G186876; JOINED.  
DR EMBL: M55348; G186876; JOINED.  
DR EMBL: M55349; G186876; JOINED.  
DR EMBL: M55350; G186876; JOINED.  
DR EMBL: M55351; G186876; JOINED.  
DR EMBL: M55352; G186876; JOINED.  
DR EMBL: M55353; G186876; JOINED.  
DR EMBL: M55355; G186876; JOINED.  
DR EMBL: M55357; G186876; JOINED.  
DR EMBL: M55357; G186876; JOINED.  
DR EMBL: M55358; G186876; JOINED.  
DR EMBL: M55359; G186876; JOINED.  
DR EMBL: M55360; G186876; JOINED.  
DR EMBL: M55361; G186876; JOINED.  
DR EMBL: M55362; G186876; JOINED.  
DR EMBL: M55363; G186876; JOINED.  
DR EMBL: M55364; G186876; JOINED.  
DR EMBL: M55366; G186876; JOINED.  
DR EMBL: M55367; G186876; JOINED.  
DR EMBL: M55368; G186876; JOINED.  
DR EMBL: M55369; G186876; JOINED.  
DR EMBL: M61916; G186837; JOINED.  
DR EMBL: M20206; G186915; -  
DR PIR: S13547; MMBU1.  
DR MIM: 150240; -  
DR PROSITE: PS00022; EGF\_1; 9.  
DR PROSITE: PS01186; EGF\_2; 2.  
DR PROSITE: PS01248; LAMININ\_TYPE\_EGF; 11.  
KW GLYCOPROTEIN; BASEMENT MEMBRANE; EXTRACELLULAR MATRIX; COILED COIL;  
LAMININ EGF-LIKE DOMAIN; CELL ADHESION; REPEAT; SIGNAL.  
FT SIGNAL 1 21  
FT CHAIN 22 1786 LAMININ BETA-1 CHAIN.  
FT DOMAIN 22 270 LAMININ N-TERMINAL (DOMAIN VI).  
FT DOMAIN 271 540 4.5 X LAMININ EGF-LIKE REPEATS (DOMAIN  
V).  
FT DOMAIN 271 334 LAMININ EGF-LIKE 1.  
FT DOMAIN 335 397 LAMININ EGF-LIKE 2.  
FT DOMAIN 398 457 LAMININ EGF-LIKE 3.  
FT DOMAIN 458 509 LAMININ EGF-LIKE 4.

FT DOMAIN 510 540 LAMININ EGF-LIKE 5 (INCOMPLETE).  
 FT DOMAIN 541 771 LAMININ DOMAIN IV.  
 FT DOMAIN 772 1178 8 X LAMININ EGF-LIKE REPEATS (DOMAIN III).  
 FT DOMAIN 773 820 LAMININ EGF-LIKE 6.  
 FT DOMAIN 821 866 LAMININ EGF-LIKE 7.  
 FT DOMAIN 867 916 LAMININ EGF-LIKE 8.  
 FT DOMAIN 917 975 LAMININ EGF-LIKE 9.  
 FT DOMAIN 976 1027 LAMININ EGF-LIKE 10.  
 FT DOMAIN 1028 1083 LAMININ EGF-LIKE 11.  
 FT DOMAIN 1084 1131 LAMININ EGF-LIKE 12.  
 FT DOMAIN 1132 1178 LAMININ EGF-LIKE 13.  
 FT DOMAIN 1179 1397 DOMAIN II.  
 FT DOMAIN 1398 1430 DOMAIN ALPHA.  
 FT DOMAIN 1431 1786 DOMAIN I.  
 FT DOMAIN 1216 1315 COILED COIL (POTENTIAL).  
 FT DOMAIN 1353 1388 COILED COIL (POTENTIAL).  
 FT DOMAIN 1442 1781 COILED COIL (POTENTIAL).  
 FT DOMAIN 271 280 BY SIMILARITY.  
 FT DISULFID 273 298 BY SIMILARITY.  
 FT DISULFID 300 309 BY SIMILARITY.  
 FT DISULFID 312 332 BY SIMILARITY.  
 FT DISULFID 335 344 BY SIMILARITY.  
 FT DISULFID 337 362 BY SIMILARITY.  
 FT DISULFID 365 374 BY SIMILARITY.  
 FT DISULFID 377 395 BY SIMILARITY.  
 FT DISULFID 398 411 BY SIMILARITY.  
 FT DISULFID 400 426 BY SIMILARITY.  
 FT DISULFID 428 437 BY SIMILARITY.  
 FT DISULFID 440 455 BY SIMILARITY.  
 FT DISULFID 458 472 BY SIMILARITY.  
 FT DISULFID 460 479 BY SIMILARITY.  
 FT DISULFID 481 490 BY SIMILARITY.  
 FT DISULFID 493 507 BY SIMILARITY.  
 FT DISULFID 773 785 BY SIMILARITY.  
 FT DISULFID 775 792 BY SIMILARITY.  
 FT DISULFID 794 803 BY SIMILARITY.  
 FT DISULFID 806 818 BY SIMILARITY.  
 FT DISULFID 821 833 BY SIMILARITY.  
 FT DISULFID 823 840 BY SIMILARITY.  
 FT DISULFID 842 851 BY SIMILARITY.  
 FT DISULFID 854 864 BY SIMILARITY.  
 FT DISULFID 867 876 BY SIMILARITY.  
 FT DISULFID 869 883 BY SIMILARITY.  
 FT DISULFID 886 895 BY SIMILARITY.  
 FT DISULFID 898 914 BY SIMILARITY.  
 FT DISULFID 917 933 BY SIMILARITY.  
 FT DISULFID 919 944 BY SIMILARITY.  
 FT DISULFID 946 955 BY SIMILARITY.  
 FT DISULFID 958 973 BY SIMILARITY.  
 FT DISULFID 976 990 BY SIMILARITY.  
 FT DISULFID 978 997 BY SIMILARITY.  
 FT DISULFID 1000 1009 BY SIMILARITY.  
 FT DISULFID 1012 1025 BY SIMILARITY.  
 FT DISULFID 1084 1096 BY SIMILARITY.  
 FT DISULFID 1106 1103 BY SIMILARITY.  
 FT DISULFID 1105 1114 BY SIMILARITY.  
 FT DISULFID 1117 1129 BY SIMILARITY.  
 FT DISULFID 1132 1144 BY SIMILARITY.  
 FT DISULFID 1134 1151 BY SIMILARITY.  
 FT DISULFID 1153 1162 BY SIMILARITY.  
 FT DISULFID 1165 1179 BY SIMILARITY.  
 FT DISULFID 1179 1179 INTERCHAIN (PROBABLE).  
 FT DISULFID 1182 1182 INTERCHAIN (PROBABLE).  
 FT DISULFID 1182 1182 INTERCHAIN (PROBABLE).  
 FT CARBOHYD 120 120 POTENTIAL.  
 FT CARBOHYD 356 356 POTENTIAL.  
 FT CARBOHYD 519 519 POTENTIAL.  
 FT CARBOHYD 677 677 POTENTIAL.  
 FT CARBOHYD 1041 1041 POTENTIAL.  
 FT CARBOHYD 1195 1195 POTENTIAL.  
 FT CARBOHYD 1279 1279 POTENTIAL.  
 FT CARBOHYD 1336 1336 POTENTIAL.

FT CARBOHYD 1343 1343 POTENTIAL.  
 FT CARBOHYD 1487 1487 POTENTIAL.  
 FT CARBOHYD 1542 1542 POTENTIAL.  
 FT CONFLICT 1470 1470 L -> V (IN REF. 3).  
 FT CONFLICT 1696 1696 E -> G (IN REF. 3).  
 SQ SEQUENCE 1786 AA; 198065 MW; 100DB223 CRC32;  
  
 Query Match 38.7%; Score 46; DB 1; Length 1786;  
 Best Local Similarity 40.0%; Pred. No. 40;  
 Matches 10; Conservative 2; Mismatches 9; Indels 4; Gaps 1;  
  
 QY 1 CNPSSGGRKVE---LVGERSIYC 21  
 Db 1105 CNPFGGRTSCSECELFWDGPDVECC 1129  
  
 RESULT 8  
 LMB1\_MOUSE STANDARD; PRT; 1786 AA.  
 ID LMB1\_MOUSE  
 AC P02469;  
 DT 21-JUL-1986 (REL. 01, CREATED)  
 DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)  
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
 DE LAMININ BETA-1 CHAIN PRECURSOR (LAMININ B1 CHAIN).  
 GN LAMB1-1 OR LAMB-1.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 87147212.  
 RA SASAKI M., KATO S., KOHNO K., MARTIN G.R., YAMADA Y.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 84:935-939(1987).  
 RN [2]  
 RP SEQUENCE OF 1292-1786 FROM N.A.  
 RX MEDLINE; 85051302.  
 RA BARLOW D.P., GREEN N.M., KURKINEN M., HOGAN B.L.M.;  
 RL EMBO J. 3:2355-2362(1984).  
 RN [3]  
 RP SEQUENCE OF 165-172; 539-547 AND 712-719.  
 RC STRAIN-BALB/C; TISSUE-ENDOTHELIAL CELLS;  
 RX MEDLINE; 97363207.  
 RA PRIESER M., NOECKEL H., PAUSCH F., ROEDER C., HAHN A., DEUTZMANN R.,  
 RA SOROKIN L.M.;  
 RL EUR. J. BIOCHEM. 246:727-735(1997).  
 CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ  
 IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF  
 CC CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING  
 CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.  
 CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE  
 CC DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND  
 CC TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE  
 CC COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.  
 CC -1- SUBUNIT: THE BETA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ),  
 CC LAMININ-2 (MEROSIN), AND LAMININ-6 (K-LAMININ).  
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
 CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR  
 CC COMPONENT).  
 CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).  
 CC -1- SIMILARITY: CONTAINS 12.5 LAMININ EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.  
 DR EMBL; M15525; G293690; ALT\_INIT.  
 DR EMBL; X05212; G809042; -.  
 DR PIR; A26413; MAMSB1.  
 DR MGD; MGI:96743; LAMB1-1.  
 DR PROSITE; PS00022; EGF\_1; 9.  
 DR PROSITE; PS01186; EGF\_2; 2.  
 DR PROSITE; PS01248; LAMININ TYPE EGF; 11.  
 KW GLYCOPROTEIN; BASEMENT MEMBRANE; EXTRACELLULAR MATRIX; COILED COIL;  
 KW LAMININ EGF-LIKE DOMAIN; CELL ADHESION; REPEAT; SIGNAL.  
 FT SIGNAL 1 21  
 FT CHAIN 22 1786 LAMININ BETA-1 CHAIN.

```

FT CARBOHYD 356 356 POTENTIAL.
FT CARBOHYD 519 519 POTENTIAL.
FT CARBOHYD 677 677 POTENTIAL.
FT CARBOHYD 1041 1041 POTENTIAL.
FT CARBOHYD 1195 1195 POTENTIAL.
FT CARBOHYD 1279 1279 POTENTIAL.
FT CARBOHYD 1336 1336 POTENTIAL.
FT CARBOHYD 1343 1343 POTENTIAL.
FT CARBOHYD 1487 1487 POTENTIAL.
FT CARBOHYD 1533 1533 POTENTIAL.
FT CARBOHYD 1542 1542 POTENTIAL.
FT CARBOHYD 1643 1643 POTENTIAL.
FT CONFLICT 1531 1534 POTENTIAL.
FT CONFLICT 1749 1749 D -> N (IN REF. 2).
SQ SEQUENCE 1786 AA; 196904 MW; 5C6A7685 CRC32;

Query Match 38.7%; Score 46; DB 1; Length 1786;
Best Local Similarity 40.0%; Pred. No. 40;
Matches 10; Conservative 2; Mismatches 9; Indels 4; Gaps 1

OY 1 CNPSSGGRKVF---LVGEPSTYC 21
| | | | | | | | | | | | | | |
Db 1105 CNPFGGRTCSECQLFWGDPDYEC 1129

RESULT 9
S24D_ANOGA STANDARD; PRF: 271 AA.
AC Q17004;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE SERINE PROTEASE SP24D PRECURSOR (EC 3.4.21.-) (AASP24D).
OS ANOPHELES GAMBIAE (AFRICAN MALARIA MOSQUITO).
OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-G3;
RX MEDLINE: 98022538.
RA HAN Y.S., SALAZAR C.E., REESE-STARDY S.R., CORNELL A., GORMAN M.J.,
RA COLLINS F.H., PASKEWITZ S.W.;
RL INSECT MOL. BIOL. 6:385-395(1997).
CC CC -1- DEVELOPMENTAL STAGE: ADULT MALES EXPRESS MUCH HIGHER LEVELS OF
CC AGSP24D mRNA THAN FEMALES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC
DR EMBL: U21917; G717150; -.
DR PROSITE: PS00134; TRYPsin_HIS. 1.
DR PROSITE: PS00135; TRYPsin_SER; FALSE NEG.
KM HYDROLASE; SERINE PROTEASE; ZYMOGEN. SIGNAL.
FT SIGNAL 1 22
FT PROPE 23 49
FT CHAIN 50 271 ACTIVATION PEPTIDE (POTENTIAL).
FT ACT_SITE 90 90 SERINE PROTEASE SP24D.
FT ACT_SITE 136 136 CHANGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 225 225 CHANGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 75 91 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 199 211 BY SIMILARITY.
FT DISULFID 221 246 BY SIMILARITY.
SQ SEQUENCE 271 AA; 28581 MW; 0B065B42 CRC32;

Query Match 38.2%; Score 45.5; DB 1; Length 271;
Best Local Similarity 42.3%; Pred. No. 6.9;
Matches 11; Conservative 3; Mismatches 7; Indels 5; Gaps 2

OY 1 CNPSSGGRKVF---LVGEPSTYC 21
| | | | | | | | | | | | | | |
Db 221 CNPDSGGPAILNNOLVGRPNFTIITC 246

RESULT 10

```



CC ATP, MAGNESIUM AND A REDUCING AGENT, SUCH AS TL(III)CITRATE, FOR  
CC ACTIVITY.  
CC -1- ENZYME REGULATION: ACTIVATED BY THE HGDC PROTEIN.  
CC -1- PATHWAY: INVOLVED IN THE 2-HYDROXY-ACID PATHWAY OF AMINO  
CC ACID FERMENTATION.  
CC -1- SUBUNIT: Tetramer of two alpha and two beta subunits.  
DR EMBL: X14252; G38800; -.  
DR PIR: S04477; DMDXAF.  
KW LYSASE; IRON-SULFUR; 4FE-4S; MAGNESIUM; FMN.  
FT INIT\_MET 0 0  
FT CONFLICT 25 25 P -> A (IN REF. 2).  
SQ SEQUENCE 476 AA; 53890 MW; DICDC158 CRC32;  
  
Query Match 37.4%; Score 44.5; DB 1; Length 476;  
Best Local Similarity 44.0%; Pred. No. 18;  
Matches 11; Conservative 3; Mismatches 4; Indels 7; Gaps 1;  
  
OY 2 NPGSG-----GRKVELVEPSI 19  
DB 112 NPHSGKMDKANGKRVEDADGKPIV 136  
  
RESULT 13  
CAH\_DUNSA STANDARD; PRT; 589 AA.  
AC P54212;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE CARBONIC ANHYDRASE (EC 4.2.1.1) (CARBONATE DEHYDRATASE).  
GN DCA.  
DN DUNALIPELLA SALINA.  
OS EUKARYOTA; PLANTA; PHYCOPHYTA; CHLOROPHYTA (GREEN ALGAE);  
OC CHLOROPHYCEAE; VOLVOCALES; DUNALIELLACEAE.  
OC [1]  
RN SEQUENCE FROM N.A.  
RP MEDLINE: 96279304.  
RA FISHER M., GOKHMAN I., PICK U., ZAMIR A.;  
RL J. BIOL. CHEM. 271:17718-17723(1996).  
CC -1- FUNCTION: REVERSIBLE HYDRATATION OF CARBON DIOXIDE.  
CC -1- CATALYTIC ACTIVITY: H(2)CO(3) -> CO(2) + H(2)O.  
CC -1- INDUCTION: BY SALT.  
CC -1- SIMILARITY: BELONGS TO THE CARBONIC ANHYDRASE FAMILY.  
DR EMBL: U53811; G1431878; -.  
KW LYSASE; ZINC.  
FT DOMAIN 390 589 CATALYTIC.  
FT METAL 418 418 ZINC (CATALYTIC) (BY SIMILARITY).  
FT METAL 420 420 ZINC (CATALYTIC) (BY SIMILARITY).  
FT METAL 440 440 ZINC (CATALYTIC) (BY SIMILARITY).  
SQ SEQUENCE 589 AA; 64257 MW; 2D375843 CRC32;  
  
Query Match 37.0%; Score 44; DB 1; Length 589;  
Best Local Similarity 66.7%; Pred. No. 26;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
  
OY 7 GRKVELVEPS 18  
DB 309 GRKVELVEPSN 320  
  
RESULT 14  
DAF1\_MOUSE STANDARD; PRT; 390 AA.  
AC Q61475; Q61397;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE COMPLEMENT DECAY-ACCELERATING FACTOR, GPI-ANCHORED PRECURSOR  
DE (DAF-GPI).  
GN DAF1.  
OS MUS MUSCULUS (MOUSE).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.  
OC [1]  
RN SEQUENCE FROM N.A.  
RP STRAIN-C57BL/6J; TISSUE-TESTIS;  
RX MEDLINE: 95403982.  
RA SPICER A.P., SELDIN M.F., GENDLER S.J.;  
RL J. IMMUNOL. 155:3079-3091(1995).  
RN [2]  
RP SEQUENCE OF 7-390 FROM N.A.  
RC STRAIN-BALB/C; TISSUE-SPLEEN;  
RX MEDLINE: 96362213.  
RA FUKUOKA Y., YASUI A., OKADA N., OKADA H.;  
RL INT. IMMUNOL. 8:379-385(1996).  
CC -1- FUNCTION: PROTECTION OF CELLS FROM COMPLEMENT-MEDIATED DAMAGE (BY  
CC SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.  
CC -1- TISSUE SPECIFICITY: BRAIN, SECRETORY EPITHELIA, SKELETAL MUSCLE,  
CC LIVER, TESTES, THYMUS, SPLEEN AND LYMPH NODE.  
CC -1- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR  
CC FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE  
CC ACTIVE SITE ON SCR3 (BY SIMILARITY).  
CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.  
CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION  
CC (RCA) FAMILY.  
DR EMBL: L41366; G886335; -.  
DR EMBL: D63679; D1010476; -.  
DR MGD: MGI:104850; DAF1.  
KW COMPLEMENT PATHWAY; GLYCOPROTEIN; MEMBRANE; REPEAT; GPI-ANCHOR;  
KW SIGNAL; SUSHI.  
FT SIGNAL 1 34 POTENTIAL.  
FT CHAIN 35 361 GPI-ANCHORED.  
FT FT PROPEP 362 390 REMOVED IN MATURE FORM (BY SIMILARITY).  
FT DOMAIN 35 285 4 X SUSHI (SCR) REPEATS.  
FT REPEAT 35 95 SUSHI 1.  
FT REPEAT 97 159 SUSHI 2.  
FT REPEAT 162 221 SUSHI 3.  
FT REPEAT 224 285 SUSHI 4.  
FT DOMAIN 288 364 SER/THR-RICH.  
FT DISULFID 65 94 BY SIMILARITY.  
FT DISULFID 98 145 BY SIMILARITY.  
FT DISULFID 129 158 BY SIMILARITY.  
FT DISULFID 163 204 BY SIMILARITY.  
FT DISULFID 190 220 BY SIMILARITY.  
FT DISULFID 225 267 BY SIMILARITY.  
FT DISULFID 253 284 BY SIMILARITY.  
FT CARBOHYD 187 187 POTENTIAL.  
FT CARBOHYD 262 262 POTENTIAL.  
FT LIPID 361 361 GPI-ANCHOR (BY SIMILARITY).  
FT CONFLICT 7 7 P -> A (IN REF. 2).  
FT CONFLICT 9 9 T -> A (IN REF. 2).  
FT CONFLICT 83 83 E -> G (IN REF. 2).  
FT CONFLICT 91 91 E -> G (IN REF. 2).  
FT CONFLICT 135 135 E -> K (IN REF. 2).  
FT CONFLICT 173 173 H -> L (IN REF. 2).  
FT CONFLICT 180 180 I -> T (IN REF. 2).  
SQ SEQUENCE 390 AA; 42618 MW; ADBBD34 CRC32;  
  
Query Match 37.0%; Score 44; DB 1; Length 390;  
Best Local Similarity 42.9%; Pred. No. 17;  
Matches 9; Conservative 2; Mismatches 4; Indels 6; Gaps 1;  
  
OY 1 CNPGSGRRVFEVLEPSIYC 21  
DB 190 CNPGSGRRVFEVLEPSIYC 204  
  
RESULT 15  
DAF2\_MOUSE STANDARD; PRT; 407 AA.  
ID DAF2\_MOUSE  
AC Q61476;

DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE COMPLEMENT DECAV-ACCELERATING FACTOR, TRANSMEMBRANE PRECURSOR  
DE (DAF-TM).  
GN DAF2.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.  
RN [1]  
RP SEQUENCE FROM N. A.  
RX STRAIN=C57BL/6J; TISSUE=TESTIS;  
RX MEDLINE; 95403962.  
RA SPICER A.P., SELDIN M.F., GENDLER S.J.;  
RL J. IMMUNOL. 155:3079-3091(1995).  
CC -1- FUNCTION: PROTECTION OF CELLS FROM COMPLEMENT-MEDIATED DAMAGE (BY  
SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).  
CC -1- TISSUE SPECIFICITY: TESTES, SPLEEN AND LYMPH NODE.  
CC -1- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR  
FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE  
ACTIVE SITE ON SCR3 (BY SIMILARITY).  
CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.  
CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION  
(RCA) FAMILY.  
DR EMBL; L41365; G886338; -.  
DR MGD; MGI:104849; DAF2.  
KW COMPLEMENT PATHWAY; GLYCOPROTEIN; REPEAT; SIGNAL; SUSHI;  
KW TRANSMEMBRANE.  
FT SIGNAL 1 39 POTENTIAL.  
FT CHAIN 40 407 COMPLEMENT DECAV-ACCELERATING FACTOR,  
TRANSMEMBRANE.  
FT DOMAIN 40 368 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 369 389 POTENTIAL.  
FT DOMAIN 390 407 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 40 290 4 X SUSHI (SCR) REPEATS.  
FT REPEAT 40 100 SUSHI 1.  
FT REPEAT 102 164 SUSHI 2.  
FT REPEAT 167 226 SUSHI 3.  
FT REPEAT 229 290 SUSHI 4.  
FT DOMAIN 291 363 SER/THR-RICH (BY SIMILARITY).  
FT DISULFID 70 99 BY SIMILARITY.  
FT DISULFID 103 150 BY SIMILARITY.  
FT DISULFID 134 163 BY SIMILARITY.  
FT DISULFID 168 209 BY SIMILARITY.  
FT DISULFID 195 225 BY SIMILARITY.  
FT DISULFID 230 272 BY SIMILARITY.  
FT DISULFID 258 289 BY SIMILARITY.  
FT CARBOHYD 192 192 POTENTIAL.  
FT CARBOHYD 267 267 POTENTIAL.  
SQ SEQUENCE 407 AA; 44469 MW; BABF07E9 CRC32;

Query Match 37.0%; Score 44; DB 1; Length 407;  
Best Local Similarity 47.6%; Pred. No. 18;  
Matches 10; Conservative 1; Mismatches 4; Indels 6; Gaps 1;

OY 1 CNPSSGRKRVFELYGEPSIYC 21  
| | | | : | | | | | |  
Db 195 CNPG-----YRLVGITISILC 209

Search completed: July 18, 1999, 11:37:51  
Job time: 1750 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OW protein - protein search, using sw model

Run on: July 18, 1999, 06:02:50 ; Search time 63.5 Seconds

(without alignments)  
18.245 Million cell updates/sec

Title: US-09-142-043-4

Perfect score: 119

Sequence: 1 CNPSSGGRRVFLVGEPSIYC 21

Scoring table: BLOSUM62

Searched: 180763 seqs, 55169189 residues

Database :

SPTREMBL\_8:\*  
1: sp\_fungi:\*  
2: sp\_human:\*  
3: sp\_invertebrate:\*  
4: sp\_mammal:\*  
5: sp\_mhc:\*  
6: sp\_organelle:\*  
7: sp\_phage:\*  
8: sp\_plant:\*  
9: sp\_bacteria:\*  
10: sp\_rodent:\*  
11: sp\_virus:\*  
12: sp\_vertibrate:\*  
13: sp\_unclassified:\*  
14: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	119	100.0	2039	2	Q16745	Q16745 homo sapien
2	119	100.0	2489	2	Q16744	Q16744 homo sapien
3	119	100.0	661	4	Q29531	Q29531 pan troglod
4	119	100.0	2014	4	Q29530	Q29530 pan troglod
5	106	89.1	315	4	Q28770	Q28770 papio hamad
6	106	89.1	1911	4	Q29528	Q29528 papio hamad
7	106	89.1	522	4	Q28797	Q28797 pan troglod
8	98	82.4	479	2	Q14079	Q14079 homo sapien
9	95	79.8	522	4	Q28769	Q28769 papio hamad
10	77	64.7	417	10	Q35520	Q35520 rattus norv
11	77	64.7	89	10	Q63129	Q63129 rattus norv
12	77	64.7	497	10	Q63612	Q63612 rattus norv
13	77	64.7	559	10	Q63135	Q63135 rattus norv
14	74	62.2	483	10	Q64735	Q64735 mus musculu
15	69	58.0	679	10	Q99254	Q99254 mus musculu
16	52	43.7	1032	2	Q13866	Q13866 homo sapien
17	52	43.7	1087	2	Q14122	Q14122 homo sapien
18	51.5	43.3	222	4	Q19120	Q19120 actus trivi
19	51.5	43.3	222	4	Q19125	Q19125 salmistr sci
20	50.5	42.4	349	2	Q15429	Q15429 homo sapien
21	50.5	42.4	336	4	Q62834	Q62834 saginus oe
22	50.5	42.4	314	4	Q62835	Q62835 saginus oe
23	50.5	42.4	378	4	Q62837	Q62837 saginus oe
24	50.5	42.4	377	4	Q62838	Q62838 saginus oe
25	50.5	42.4	222	4	Q19122	Q19122 callimico g
26	50.5	42.4	222	4	Q19123	Q19123 callithrix
27	50.5	42.4	222	4	Q19124	Q19124 saginus my
28	50.5	42.4	222	4	Q19128	Q19128 pithecia pi
29	50	42.0	347	2	P78361	P78361 homo sapien

## ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	2039 AA.
Q16745	Q16745	Q16745	Q16745	Q16745
AC	Q16745	Q16745	Q16745	Q16745
DT	01-NOV-1996 (TREMBLREL. 01, CREATED)	01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)	01-NOV-1996 (TREMBLREL. 08, LAST ANNOTATION UPDATE)	01-NOV-1996 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE	COMPLEMENT RECEPTOR 1.	COMPLEMENT RECEPTOR 1.	COMPLEMENT RECEPTOR 1.	COMPLEMENT RECEPTOR 1.
OS	HOMO SAPIENS (HUMAN).	HOMO SAPIENS (HUMAN).	HOMO SAPIENS (HUMAN).	HOMO SAPIENS (HUMAN).
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
CA	CATARRHINI; HOMINIDAE; HOMO.	CATARRHINI; HOMINIDAE; HOMO.	CATARRHINI; HOMINIDAE; HOMO.	CATARRHINI; HOMINIDAE; HOMO.
RN	[1]	[1]	[1]	[1]
RP	SEQUENCE FROM N.A.	SEQUENCE FROM N.A.	SEQUENCE FROM N.A.	SEQUENCE FROM N.A.
RX	MEDLINE: 94065175.	MEDLINE: 94065175.	MEDLINE: 94065175.	MEDLINE: 94065175.
RA	VIR D.P., WONG W.W.;	VIR D.P., WONG W.W.;	VIR D.P., WONG W.W.;	VIR D.P., WONG W.W.;
RT	"Structure of the gene for the F allele of complement receptor type 1 and sequence of the coding region unique to the S allele."	"Structure of the gene for the F allele of complement receptor type 1 and sequence of the coding region unique to the S allele."	"Structure of the gene for the F allele of complement receptor type 1 and sequence of the coding region unique to the S allele."	"Structure of the gene for the F allele of complement receptor type 1 and sequence of the coding region unique to the S allele."
RJ	J. IMMUNOL. 151:6214-6224(1993).	J. IMMUNOL. 151:6214-6224(1993).	J. IMMUNOL. 151:6214-6224(1993).	J. IMMUNOL. 151:6214-6224(1993).
RN	[2]	[2]	[2]	[2]
RP	SEQUENCE FROM N.A.	SEQUENCE FROM N.A.	SEQUENCE FROM N.A.	SEQUENCE FROM N.A.
RA	VIR D.P., WONG W.W.;	VIR D.P., WONG W.W.;	VIR D.P., WONG W.W.;	VIR D.P., WONG W.W.;
RT	SUBMITTED (JUN-1993) TO EMBL/GENBANK/DBJ DATA BANKS.	SUBMITTED (JUN-1993) TO EMBL/GENBANK/DBJ DATA BANKS.	SUBMITTED (JUN-1993) TO EMBL/GENBANK/DBJ DATA BANKS.	SUBMITTED (JUN-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
DR	EMBL: L17399; G306680; JOINED.	EMBL: L17399; G306680; JOINED.	EMBL: L17399; G306680; JOINED.	EMBL: L17399; G306680; JOINED.
DR	EMBL: L17409; G306680; JOINED.	EMBL: L17409; G306680; JOINED.	EMBL: L17409; G306680; JOINED.	EMBL: L17409; G306680; JOINED.
DR	EMBL: L17420; G306680; JOINED.	EMBL: L17420; G306680; JOINED.	EMBL: L17420; G306680; JOINED.	EMBL: L17420; G306680; JOINED.
DR	EMBL: L17421; G306680; JOINED.	EMBL: L17421; G306680; JOINED.	EMBL: L17421; G306680; JOINED.	EMBL: L17421; G306680; JOINED.
DR	EMBL: L17422; G306680; JOINED.	EMBL: L17422; G306680; JOINED.	EMBL: L17422; G306680; JOINED.	EMBL: L17422; G306680; JOINED.
DR	EMBL: L17423; G306680; JOINED.	EMBL: L17423; G306680; JOINED.	EMBL: L17423; G306680; JOINED.	EMBL: L17423; G306680; JOINED.
DR	EMBL: L17391; G306680; JOINED.	EMBL: L17391; G306680; JOINED.	EMBL: L17391; G306680; JOINED.	EMBL: L17391; G306680; JOINED.
DR	EMBL: L17382; G306680; JOINED.	EMBL: L17382; G306680; JOINED.	EMBL: L17382; G306680; JOINED.	EMBL: L17382; G306680; JOINED.
DR	EMBL: L17393; G306680; JOINED.	EMBL: L17393; G306680; JOINED.	EMBL: L17393; G306680; JOINED.	EMBL: L17393; G306680; JOINED.
DR	EMBL: L17394; G306680; JOINED.	EMBL: L17394; G306680; JOINED.	EMBL: L17394; G306680; JOINED.	EMBL: L17394; G306680; JOINED.
DR	EMBL: L17395; G306680; JOINED.	EMBL: L17395; G306680; JOINED.	EMBL: L17395; G306680; JOINED.	EMBL: L17395; G306680; JOINED.
DR	EMBL: L17396; G306680; JOINED.	EMBL: L17396; G306680; JOINED.	EMBL: L17396; G306680; JOINED.	EMBL: L17396; G306680; JOINED.
DR	EMBL: L17397; G306680; JOINED.	EMBL: L17397; G306680; JOINED.	EMBL: L17397; G306680; JOINED.	EMBL: L17397; G306680; JOINED.
DR	EMBL: L17398; G306680; JOINED.	EMBL: L17398; G306680; JOINED.	EMBL: L17398; G306680; JOINED.	EMBL: L17398; G306680; JOINED.
DR	EMBL: L17418; G306680; JOINED.	EMBL: L17418; G306680; JOINED.	EMBL: L17418; G306680; JOINED.	EMBL: L17418; G306680; JOINED.
DR	EMBL: L17390; G306680; JOINED.	EMBL: L17390; G306680; JOINED.	EMBL: L17390; G306680; JOINED.	EMBL: L17390; G306680; JOINED.
DR	EMBL: L17400; G306680; JOINED.	EMBL: L17400; G306680; JOINED.	EMBL: L17400; G306680; JOINED.	EMBL: L17400; G306680; JOINED.
DR	EMBL: L17416; G306680; JOINED.	EMBL: L17416; G306680; JOINED.	EMBL: L17416; G306680; JOINED.	EMBL: L17416; G306680; JOINED.
DR	EMBL: L17417; G306680; JOINED.	EMBL: L17417; G306680; JOINED.	EMBL: L17417; G306680; JOINED.	EMBL: L17417; G306680; JOINED.
DR	EMBL: L17401; G306680; JOINED.	EMBL: L17401; G306680; JOINED.	EMBL: L17401; G306680; JOINED.	EMBL: L17401; G306680; JOINED.
DR	EMBL: L17402; G306680; JOINED.	EMBL: L17402; G306680; JOINED.	EMBL: L17402; G306680; JOINED.	EMBL: L17402; G306680; JOINED.
DR	EMBL: L17403; G306680; JOINED.	EMBL: L17403; G306680; JOINED.	EMBL: L17403; G306680; JOINED.	EMBL: L17403; G306680; JOINED.
DR	EMBL: L17404; G306680; JOINED.	EMBL: L17404; G306680; JOINED.	EMBL: L17404; G306680; JOINED.	EMBL: L17404; G306680; JOINED.
DR	EMBL: L17405; G306680; JOINED.	EMBL: L17405; G306680; JOINED.	EMBL: L17405; G306680; JOINED.	EMBL: L17405; G306680; JOINED.
DR	EMBL: L17406; G306680; JOINED.	EMBL: L17406; G306680; JOINED.	EMBL: L17406; G306680; JOINED.	EMBL: L17406; G306680; JOINED.
DR	EMBL: L17407; G306680; JOINED.	EMBL: L17407; G306680; JOINED.	EMBL: L17407; G306680; JOINED.	EMBL: L17407; G306680; JOINED.
DR	EMBL: L17408; G306680; JOINED.	EMBL: L17408; G306680; JOINED.	EMBL: L17408; G306680; JOINED.	EMBL: L17408; G306680; JOINED.
DR	EMBL: L17410; G306680; JOINED.	EMBL: L17410; G306680; JOINED.	EMBL: L17410; G306680; JOINED.	EMBL: L17410; G306680; JOINED.
DR	EMBL: L17411; G306680; JOINED.	EMBL: L17411; G306680; JOINED.	EMBL: L17411; G306680; JOINED.	EMBL: L17411; G306680; JOINED.

DR EMBL; L17412; G306680; JOINED.  
DR EMBL; L17413; G306680; JOINED.  
DR EMBL; L17414; G306680; JOINED.  
DR EMBL; L17415; G306680; JOINED.  
DR PFAM; PF00084; sushi; 30  
SQ SEQUENCE 2039 AA; 223603 MW; A4972215 CRC32;

Query Match 100.0%; Score 119; DB 2; Length 2039;  
Best Local Similarity 100.0%; Pred. No. 6, 2e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVFELVGEPSIYC 21  
|||||  
Db 195 CNPSSGGRKVFELVGEPSIYC 215

RESULT 2  
ID Q16744 PRELIMINARY; PRT; 2489 AA.  
AC Q16744;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE COMPLEMENT RECEPTOR 1.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; HOMO.  
RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE; 94065175.  
RA VIK D.P., WONG W.W.;  
RT "Structure of the gene for the F allele of complement receptor type 1  
J. IMMUNOL. 151:6214-6224(1993)."  
RL [2]

RN SEQUENCE FROM N.A.  
RA VIK D.P., WONG W.W.;  
RL SUBMITTED (JUN-1993) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL; L17399; G451303; JOINED.  
DR EMBL; L17409; G451303; JOINED.  
DR EMBL; L17419; G451303; JOINED.  
DR EMBL; L17420; G451303; JOINED.  
DR EMBL; L17421; G451303; JOINED.  
DR EMBL; L17422; G451303; JOINED.  
DR EMBL; L17423; G451303; JOINED.  
DR EMBL; L17424; G451303; JOINED.  
DR EMBL; L17425; G451303; JOINED.  
DR EMBL; L17426; G451303; JOINED.  
DR EMBL; L17427; G451303; JOINED.  
DR EMBL; L17428; G451303; JOINED.  
DR EMBL; L17429; G451303; JOINED.  
DR EMBL; L17430; G451303; JOINED.  
DR EMBL; L17391; G451303; JOINED.  
DR EMBL; L17392; G451303; JOINED.  
DR EMBL; L17393; G451303; JOINED.  
DR EMBL; L17394; G451303; JOINED.  
DR EMBL; L17418; G451303; JOINED.  
DR EMBL; L17390; G451303; JOINED.  
DR EMBL; L17395; G451303; JOINED.  
DR EMBL; L17416; G451303; JOINED.  
DR EMBL; L17417; G451303; JOINED.  
DR EMBL; L17396; G451303; JOINED.  
DR EMBL; L17397; G451303; JOINED.  
DR EMBL; L17398; G451303; JOINED.  
DR EMBL; L17400; G451303; JOINED.  
DR EMBL; L17401; G451303; JOINED.  
DR EMBL; L17402; G451303; JOINED.  
DR EMBL; L17403; G451303; JOINED.  
DR EMBL; L17405; G451303; JOINED.  
DR EMBL; L17406; G451303; JOINED.  
DR EMBL; L17407; G451303; JOINED.  
DR EMBL; L17408; G451303; JOINED.

DR EMBL; L17410; G451303; JOINED.  
DR EMBL; L17411; G451303; JOINED.  
DR EMBL; L17412; G451303; JOINED.  
DR EMBL; L17413; G451303; JOINED.  
DR EMBL; L17414; G451303; JOINED.  
DR EMBL; L17415; G451303; JOINED.  
DR PFAM; PF00084; sushi; 37.  
SQ SEQUENCE 2489 AA; 272846 MW; 5869B6F9 CRC32;

Query Match 100.0%; Score 119; DB 2; Length 2489;  
Best Local Similarity 100.0%; Pred. No. 7, 6e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVFELVGEPSIYC 21  
|||||  
Db 195 CNPSSGGRKVFELVGEPSIYC 215

RESULT 3  
ID Q29531 PRELIMINARY; PRT; 661 AA.  
AC Q29531;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE COMPLEMENT RECEPTOR 1 PRECURSOR (FRAGMENT).  
GN CRL.  
OS PAN TROGLODYTES (CHIMPANZEE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; PAN.  
RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE; 94292799.  
RA BIRNINGHAM D.J., SHEN X.P., HOURCADE D., NICKELLS M.W., ATKINSON J.P.;  
RT "Primary sequence of an alternatively spliced form of CRL. Candidate  
for the 75,000 M(r) complement receptor expressed on chimpanzee  
erythrocytes."  
RL J. IMMUNOL. 153:691-700(1994).  
DR EMBL; L24921; G557727; -  
DR PFAM; PF00084; sushi; 9.

KW SIGNAL; ALTERNATIVE SPLICING.  
FT NON\_TER 1 1  
FT SIGNAL 16  
FT CHAIN 17 >661 POTENTIAL.  
FT NON\_TER 661 661 COMPLEMENT RECEPTOR 1.  
SQ SEQUENCE 661 AA; 72966 MW; 9D78E262 CRC32;

Query Match 100.0%; Score 119; DB 4; Length 661;  
Best Local Similarity 100.0%; Pred. No. 2, 1e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVFELVGEPSIYC 21  
|||||  
Db 170 CNPSSGGRKVFELVGEPSIYC 190

RESULT 4  
ID Q29530 PRELIMINARY; PRT; 2014 AA.  
AC Q29530;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE COMPLEMENT RECEPTOR 1 (FRAGMENT).  
GN CRL.  
OS PAN TROGLODYTES (CHIMPANZEE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; PAN.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 94292799.





DR EMBL; M31235; G563326; JOINED.  
 DR EMBL; M31236; G563326; JOINED.  
 DR PFAM; PF00084; sushi; 7.  
 FT NON\_TER 479  
 SQ SEQUENCE 479 AA; 52533 MW; EC994E0D CRC32;

Query Match 82.4%; Score 98; DB 2; Length 479;  
 Best Local Similarity 85.7%; Pred. No. 2.5e-07;  
 Matches 18; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVELVGEPSIYC 21  
 |||:|||||  
 Db 195 CNLSSGRKVELVGEPSIYC 215

RESULT 9  
 Q28769 PRELIMINARY; PRT; 522 AA.

AC Q28769;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE COMPLEMENT RECEPTOR (FRAGMENT).  
 OS PAPIO HAMADRYAS (HAMADRYAS BABOON).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
 OC CATARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; PAPIO.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BONE MARROW;  
 RA BIRMINGHAM D.J.; LOGAR C.M.; SHEN X.P.; CHEN W.;  
 RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; L77977; G1301609; -  
 DR PFAM; PF00084; sushi; 7.  
 FT NON\_TER 1  
 SQ SEQUENCE 522 AA; 56626 MW; 7862072C CRC32;

Query Match 79.8%; Score 95; DB 4; Length 522;  
 Best Local Similarity 81.0%; Pred. No. 7.9e-07;  
 Matches 17; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVELVGEPSIYC 21  
 |||:|||||  
 Db 200 CNLSSGRKVELVGEPSIYC 220

RESULT 10  
 Q35520 PRELIMINARY; PRT; 417 AA.

AC Q35520;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE 512 ANTIGEN (FRAGMENT).  
 OS RATTUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCUROGNATHI; MORIDAE; MORINAE; RATTUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LIVER;  
 RA DOHI N.; SAKURADA C.; NONAKA M.; OKADA N.; OKADA H.;  
 RL SUBMITTED (NOV-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; D42116; D1023417; -  
 DR PFAM; PF00084; sushi; 6.  
 FT NON\_TER 417  
 SQ SEQUENCE 417 AA; 45951 MW; A6F919B4 CRC32;

Query Match 64.7%; Score 77; DB 10; Length 417;  
 Best Local Similarity 61.9%; Pred. No. 0.00036;  
 Matches 13; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVELVGEPSIYC 21  
 |||:|||||  
 Db 52 CNTDARGKLFNLVGEPSIHC 72

RESULT 11  
 Q63129 PRELIMINARY; PRT; 89 AA.

AC Q63129;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE COMPLEMENT RECEPTOR TYPE 1 (FRAGMENT).  
 GN CR1.  
 OS RATTUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCUROGNATHI; MORIDAE; MORINAE; RATTUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY;  
 RX MEDLINE; 95002704.  
 RA CINGG R.J.; SNEED A.E.;  
 RT "Molecular characterization of rat glomerular epithelial cell  
 complement receptors."  
 RL J. AM. SOC. NEPHROL. 4:1912-1919(1994).  
 DR EMBL; L19118; G682650; -  
 DR PFAM; PF00084; sushi; 1.  
 KW TRANSMEMBRANE.  
 FT NON\_TER 1  
 FT NON\_TER 89  
 SQ SEQUENCE 89 AA; 9758 MW; 9FBA44EB CRC32;

Query Match 64.7%; Score 77; DB 10; Length 89;  
 Best Local Similarity 61.9%; Pred. No. 8.1e-05;  
 Matches 13; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 CNPSSGGRKVELVGEPSIYC 21  
 |||:|||||  
 Db 62 CNTDARGKLFNLVGEPSIHC 82

RESULT 12  
 Q63612 PRELIMINARY; PRT; 497 AA.

AC Q63612;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE 512 ANTIGEN PRECURSOR.  
 OS RATTUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCUROGNATHI; MORIDAE; MORINAE; RATTUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LIVER;  
 RX MEDLINE; 94161746.  
 RA SAKURADA C.; SENO H.; DOHI N.; TAKIZAWA H.; NONAKA M.; OKADA N.;  
 RA OKADA H.;  
 RT "Molecular cloning of the rat complement regulatory protein, 512  
 antigen."  
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 198:819-826(1994).  
 DR EMBL; D42114; G987688; -  
 DR PFAM; PF00084; sushi; 6.  
 KW SIGNAL.  
 FT SIGNAL 1  
 FT CHAIN 37  
 FT CHAIN 497  
 SQ SEQUENCE 497 AA; 54786 MW; 076547C7 CRC32;

Query Match 64.7%; Score 77; DB 10; Length 497;  
 Best Local Similarity 61.9%; Pred. No. 0.00043;  
 Matches 13; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 CNPSSGGRKVELVEPSIYC 21  
 ID 063135 PRELIMINARY; PRT; 559 AA.  
 AC 063135;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE COMPLEMENT REGULATORY PROTEIN.  
 GN CRRY.  
 OS RATTUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY;  
 RX MEDLINE: 96006570.  
 RA QUING R.J., LO C.F., ALEXANDER J.J., SNEED A.E., MOXLEY G. III;  
 RT "Molecular characterization of rat Crry: widespread distribution of  
 two alternative forms of Crry mRNA."  
 RL IMMUNOGENETICS 42:362-367(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LIVER.  
 RA DOHI N., SAKURADA C., NONAKA M., OKADA N., OKADA H.;  
 RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: L36532; G1202095; -  
 DR EMBL: DA2115; D1023416; -  
 DR PFAM: PF00084; sushi; 7  
 SQ SEQUENCE 559 AA; 61680 MW; C87ECC58 CRC32;

Query Match 64.7%; Score 77; DB 10; Length 559;  
 Best Local Similarity 61.9%; Pred. No. 0.00048;  
 Matches 13; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 CNPSSGGRKVELVEPSIYC 21  
 ID 064735 PRELIMINARY; PRT; 483 AA.  
 AC 064735;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE COMPLEMENT RECEPTOR RELATED PROTEIN.  
 GN CRRY.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C;  
 RX MEDLINE: 90171600.  
 RA PAULI M.S., AEGERTER-SHAW M., CEPPEK K., MILLER M.D., WEIS J.H.;  
 RT "The murine complement receptor gene family. III. The genomic and  
 transcriptional complexity of the Crry and Crry-ps genes."  
 RL J. IMMUNOL. 144:1988-1996(1990).  
 DR EMBL: M34166; G387133; JOINED.  
 DR EMBL: M34167; G387133; JOINED.  
 DR EMBL: M34168; G387133; JOINED.  
 DR EMBL: M34169; G387133; JOINED.  
 DR EMBL: M34170; G387133; JOINED.  
 DR EMBL: M34171; G387133; JOINED.  
 DR EMBL: M34172; G387133; JOINED.

DR EMBL: M34173; G387132; -  
 DR EMBL: M34173; G387133; -  
 DR EMBL: M34164; G387133; JOINED.  
 DR EMBL: M34165; G387133; JOINED.  
 DR EMBL: M34164; G387132; JOINED.  
 DR EMBL: M34170; G387132; JOINED.  
 DR EMBL: M34171; G387132; JOINED.  
 DR EMBL: M34172; G387132; JOINED.  
 DR EMBL: M34165; G387132; JOINED.  
 DR EMBL: M34166; G387132; JOINED.  
 DR EMBL: M34167; G387132; JOINED.  
 DR EMBL: M34168; G387132; JOINED.  
 DR EMBL: M34169; G387132; JOINED.  
 DR MGI: 88513; CRRY.  
 DR PFAM: PF00084; sushi; 5.  
 FT VARIANT 41 83 MISSING (IN LIVER ISOFORM).  
 SQ SEQUENCE 483 AA; 53762 MW; 70C73A80 CRC32;

Query Match 62.2%; Score 74; DB 10; Length 483;  
 Best Local Similarity 57.1%; Pred. No. 0.0012;  
 Matches 12; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 CNPSSGGRKVELVEPSIYC 21  
 ID 099254 PRELIMINARY; PRT; 679 AA.  
 AC 099254;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE COMPLEMENT RECEPTOR TYPE 2 PRECURSOR (CR2)  
 DE (ALTERNATIVELY SPLICED VERSION) (FRAGMENT).  
 GN CR2.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C;  
 RX MEDLINE: 90229754.  
 RA KURTZ C.B., O'TOOLE E., CHRISTENSEN S.M., WEIS J.H.;  
 RT "The murine complement receptor gene family. IV. Alternative splicing  
 of Cr2 gene transcripts predicts two distinct gene products that share  
 homologous domains with both human CR2 and CR1."  
 RL J. IMMUNOL. 144:3581-3591(1990).  
 RN [2]  
 RP SEQUENCE OF 21-367 FROM N.A.  
 RX MEDLINE: 95105691.  
 RA KIM Y.U., KINOSHITA T., MOLINA H., HOURCADE D., SEVA T., WAGNER L.M.,  
 RA HOLERS V.M.;  
 RT "Mouse complement regulatory protein Crry/p65 uses the specific  
 mechanisms of both human decay-accelerating factor and membrane  
 cofactor protein."  
 RL J. EXP. MED. 181:151-159(1995).  
 CC -1- TISSUE SPECIFICITY: B LYMPHOCYTES.  
 CC -1- ALTERNATIVE PRODUCTS: ALTERNATIVE SPLICING OF CR2 GENE  
 TRANSCRIPTS PREDICTS TWO DISTINCT GENE PRODUCTS THAT SHARE  
 HOMOLOGOUS DOMAINS WITH BOTH HUMAN CR2 AND CR1.  
 CC EMBL: U17124; G595982; JOINED.  
 DR EMBL: U17128; G595982; -  
 DR EMBL: U17123; G595982; JOINED.  
 DR EMBL: M36470; G192691; -  
 DR EMBL: U17127; G595982; JOINED.  
 DR EMBL: U17126; G595982; JOINED.  
 DR EMBL: U17125; G595982; JOINED.  
 DR MGI: 88489; CR2.  
 DR PFAM: PF00084; sushi; 10.  
 KW COMPLEMENT PATHWAY; GLYCOPROTEIN; TRANSMEMBRANE; REPEAT; SIGNAL;



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: July 18, 1999, 00:42:53 ; Search time 58.51 Seconds  
(without alignments)  
5.876 Million cell updates/sec

Title: US-09-142-043-5

Perfect score: 91

Sequence: 1 SSGRKVFELVGEPSIYC 17

Scoring table: BLOSUM62

Searched: 162890 seqs, 20225328 residues

Database : A\_Geneseq\_34:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	91	100.0	2317	1 P92219	CRI protein New nu
2	91	100.0	2039	1 R11810	Human complement t
3	91	100.0	1537	1 R11982	Partial human comp
4	91	100.0	543	1 R28543	CRI-4 (35E, 37Y) a
5	91	100.0	543	1 R28544	CRI-4 (35E) analog
6	91	100.0	543	1 R28545	CRI-4 (37Y) analog
7	91	100.0	543	1 R28546	CRI-4 (44T, 47D, 4
8	91	100.0	543	1 R28547	CRI-4 (52S, 53S, 5
9	91	100.0	543	1 R28548	CRI-4 (57V, 59K) a
10	91	100.0	543	1 R28549	CRI-4 (64K, 65T) a
11	91	100.0	543	1 R28550	CRI-4 (64K) analog
12	91	100.0	543	1 R28551	CRI-4 (65T) analog
13	91	100.0	543	1 R28552	CRI-4 (78T, 79D) a
14	91	100.0	543	1 R28553	CRI-4 (85R, 87N) a
15	91	100.0	543	1 R28554	CRI-4 (92T, 94H) a
16	91	100.0	543	1 R28555	CRI-4 (92T) analog
17	91	100.0	543	1 R28556	CRI-4 (94H) analog
18	91	100.0	543	1 R28557	CRI-4 (99H, 103E)
19	91	100.0	543	1 R28558	CRI-4 (109N, 110A,
20	91	100.0	543	1 R28559	CRI-4 (114-117STRP
21	91	100.0	543	1 R28560	CRI-4 (114S) analo
22	91	100.0	543	1 R28561	CRI-4 (115T) analo
23	91	100.0	543	1 R28562	CRI-4 (116K) analo
24	91	100.0	543	1 R28563	CRI-4 (117P) analo
25	91	100.0	543	1 R28564	CRI-4 (116K, 117P)
26	91	100.0	543	1 R28565	CRI-4 (1210) analo
27	91	100.0	483	1 R29091	CRI-4 (amino acids
28	91	100.0	481	1 R29092	CRI-4 (amino acids
29	91	100.0	543	1 R28566	CRI-4 (318R, 319N)
30	91	100.0	543	1 R28567	CRI-4 (318-321 RNP
31	91	100.0	543	1 R28568	CRI-4 (347T, 349Y)
32	91	100.0	543	1 R28569	CRI-4 (369-376 STR
33	91	100.0	543	1 R28570	CRI-4 (266-274 KIK
34	91	100.0	543	1 R28571	CRI-4 (364-367 NAK
35	91	100.0	2039	1 R36743	CRI. Nucleic acid
36	91	100.0	197	1 R47152	Sequence of solubl
37	91	100.0	76	1 R47153	Sequence of solubl
38	91	100.0	254	1 R47154	Sequence of solubl
39	91	100.0	254	1 R47155	Sequence of solubl
40	91	100.0	133	1 R47156	Sequence of solubl
41	91	100.0	17	1 W31836	Peptide 2 from the
42	91	100.0	21	1 W31835	Peptide 1 from the
43	91	100.0	211	1 W45909	SCR 1-3 of complem

## ALIGNMENTS

44 91 100.0 209 1 W45912  
45 91 100.0 1930 1 W45899

SCR 1-3 of complem  
Human complement r

```
RESULT 1
P92219
ID P92219 standard; protein: 2317 AA.
AC P92219;
DE 22-FEB-1990 (first entry)
DE CRI protein
OS Complement; cofactor.
OS Homo sapiens (human).
FH Key location/Qualifiers
FT peptide 10..50
FT /label= signal_peptide
PN W08909220-A.
PD 05-OCT-1989.
PE 31-MAR-1989: U01358.
PR 01-APR-1988; US-176532.
PA (TCEL) T Cell Sciences Inc; (UJO) The Johns Hopkins University;
PA (BRIG*) The Brigham and Women's Hospital.
PI Fearon DT, Klickstein LB, Wong W, Carson G, Concino MF, Makrides SC;
PI WPI: 89-309498/42.
DR N-PSDB: N91477.
PT New nucleic acid sequences encoding new CRI protein - and its fragment,
PT for diagnosis and control of complement-related immune defects,
PT inflammation, myocardial infarct, etc
PS Claim 1, fig. 1; 191pp; English.
SC This is full-length CRI protein, and shortened forms are new, lacking
CC the transmembrane region. The proteins and fragments bind C3b and/or
CC C4b, have cofactor activity and inhibit C3 and C5 convertase activity.
CC In the sequence, x=untranslated region. This has 7 short consensus
CC repeats (CR) of 4 long direct homologous repeats (LHR) (see fig. 10 in
CC patent). Deletion mutants, eg LHR-A for binding of C4b, and LHR-C
CC for C3b binding, be made, and LHR-B and -C can detect CR2 sequences.
CC They are useful in diagnosing and treating immune disorders, and prevent
CC perusion injury.
SQ Sequence 2317 AA:

Query Match 100.0%; Score 91; DB 1; Length 2317;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSGRKVFELVGEPSIYC 17
DB 208 SSGRKVFELVGEPSIYC 224

RESULT 2
R11810
ID R11810 standard; Protein: 2039 AA.
AC R11810;
DE 25-JUN-1991 (first entry)
DE Human complement type 1 receptor.
DE Complement system; C3b/C4b receptor; CRI; allergic reaction;
KW Immune response; clone lambda T109.1.
OS Homo sapiens.
FH Key location/Qualifiers
FT peptide 1..41
FT /label= putative signal peptide
FT protein 42..2039
FT /label= CRI
PN W09105047-A.
PD 18-APR-1991.
PE 25-SEP-1990; U05454.
PR 26-SEP-1989; US-412745.
PR 26-SEP-1990; US-912349.
PA (TCEL-) T CELL SCI INC.
PA (UJO) JOHNS HOPKINS UNIVERSITY.
```

PA (BRIG-) BRIGHAM AND WOMEN'S HOSPITAL.  
 PI Fearon DT, Klickstein LB, Wong WM, Carson GR, Hoh M, Conclino MF,  
 PI Makrides SC, Marsh HC;  
 DR WPI: 91-132854/18.  
 DR N-PSDB: 011642.  
 PT Human complement receptor type 1 gene, encoded proteins and  
 PT fragments - for treatment of immune disorders, myocardial infarct,  
 PT damage due to inflammation and in treatment of thrombosis  
 PS Claim 41: Fig 1: 234pp; English.  
 CC The invention also covers fragments of this protein which have the  
 CC ability to bind C3b and/or C4b, have cofactor I activity or can  
 CC inhibit C3 or C5 convertase activity. The full-length protein, or  
 CC its specified fragments are used to treat patients with immune  
 CC disorders or a disorder caused by inappropriate complement  
 CC activity. The protein is also used to treat thrombotic conditions  
 CC in humans and animals. See also 011643.  
 SO Sequence 2039 AA;

Query Match 100.0%; Score 91; DB 1; Length 2039;  
 Best Local Similarity 100.0%; Pred. No. 9.8e-08;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SGGKRVFELVGEPSIYC 17  
 |||||||  
 DB 199 SGGKRVFELVGEPSIYC 215

RESULT 3  
 R11982  
 ID R11982 standard; Protein; 1537 AA.  
 AC R11982;  
 DT 25-JUN-1991 (first entry)  
 DE Partial human complement type 1 receptor.  
 KW Complement system; C3b/C4b receptor; CRI; allergic reaction;  
 KM Immune response; long homologous repeat; LHR.  
 OS Homo sapiens.  
 FH Key  
 FT region  
 FT 1..438  
 FT /label= LHR-B  
 FT 439..891  
 FT /label= LHR-C  
 FT 892..1341  
 FT /label= LHR-D  
 FT 1495..1498  
 FT /note= "positively-charged; preceded by hydrophobic  
 FT sequence"  
 FT 1521..1526  
 FT /note= "has 67 per cent homology to site of protein  
 FT kinase C phosphorylation in the EGF  
 FT receptor"

region  
 WO9105047-A.  
 PD 18-APR-1991.  
 PD 25-SEP-1990; US-05454.  
 PR 26-SEP-1989; US-412745.  
 PR 26-SEP-1990; US-912349.  
 PA (TCELL-) T CELL SCI INC.  
 PA (UYJO ) JOHNS HOPKINS UNIVERSITY.  
 PA (BRIG-) BRIGHAM AND WOMEN'S HOSPITAL.  
 PI Fearon DT, Klickstein LB, Wong WM, Carson GR, Hoh M, Conclino MF,  
 PI Makrides SC, Marsh HC;  
 DR WPI: 91-132854/18.  
 DR N-PSDB: 011643.  
 PT Human complement receptor type 1 gene, encoded proteins and  
 PT fragments - for treatment of immune disorders, myocardial infarct,  
 PT damage due to inflammation and in treatment of thrombosis  
 PS Disclosure; Fig 5; 234pp; English.  
 CC This sequence comprises three of the four tandem, direct, long  
 CC homologous repeats of the full-length F allozyme of CRI. LHR-A is  
 CC absent. Each LHR might represent a single C3b/C4b binding domain,  
 CC making the receptor multivalent. The LHR's are composed of 7 short  
 CC consensus repeats of 60-70 residues resembling the SCR's of other  
 CC C3/C4 binding proteins. The protein and fragments of it having C3b

CC and/or C4b binding activity can be used to treat immune disorders  
 CC or disorders involving inappropriate complement activity.  
 CC See also 011642.  
 SO Sequence 1537 AA;

Query Match 100.0%; Score 91; DB 1; Length 1537;  
 Best Local Similarity 100.0%; Pred. No. 7.2e-08;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SGGKRVFELVGEPSIYC 17  
 |||||||  
 DB 147 SGGKRVFELVGEPSIYC 163

RESULT 4  
 R28543  
 ID R28543 standard; peptide; 543 AA.  
 AC R28543;  
 DT 19-MAR-1993 (first entry)  
 DE CRI-4 (35E, 37Y) analogue.  
 KW Short consensus repeat; regulator of complement activation;  
 KM C3b binding; C4b binding; human complement type 1 receptor.  
 OS Homo sapiens.  
 FH Key  
 FT region  
 FT 1..60  
 FT /label= SCR-1  
 FT 61..122  
 FT /label= SCR-2  
 FT 451..510  
 FT /label= SCR-8  
 FT 511..543  
 FT /label= SCR-9  
 FT /note= "TRUNCATED"

FT misc\_difference 35  
 FT /note= "gly substituted by glu (SCR-8)"  
 FT misc\_difference 37  
 FT /note= "Ser substituted by Tyr (SCR-8)"

FT EP-512733-A.  
 PD 11-NOV-1992.  
 PD 28-APR-1992; 303826.  
 PR 03-MAY-1991; US-695514.  
 PA (UNIW ) UNIV WASHINGTON.  
 PI Atkinson JP, Hourcade D, Krych M;  
 DR WPI: 92-375009/46.  
 PT Complement activity regulator protein analogues - useful for  
 PT treating auto-immune diseases, to suppress transplant rejection,  
 PT for diagnosis etc.  
 PS Claim 11: Fig 2 and R11810; 23pp; English.  
 CC The cDNA clone designated CRI-4 was described in J. Exp. Med. (1988)  
 CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
 CC SCRs of CRI. The invention concerns analogues of "regulator of  
 CC complement activation" proteins or truncated, hybrid or recombinant  
 CC forms of them. CRI-4 is a preferred truncated form and a number of  
 CC specified substitution variants of it are claimed. Positions 35 and  
 CC 37 of SCR-1 and the corresponding positions in SCR-8 have been  
 CC identified as important in C4b binding. The specification does not  
 CC contain the CRI-4 sequence; the sequence given here was constructed  
 CC from the full-length CRI amino acid sequence having GENESEQ  
 CC accession number R11810 and descriptions in the disclosure.  
 SO Sequence 543 AA;

Query Match 100.0%; Score 91; DB 1; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-08;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SGGKRVFELVGEPSIYC 17  
 |||||||  
 DB 158 SGGKRVFELVGEPSIYC 174

RESULT 5

```

R28544
ID R28544 standard; peptide: 543 AA.
AC R28544;
DT 19-MAR-1993 (first entry)
DE CRI-4 (35E) analogue.
KW short consensus repeat; regulator of complement activation;
  C3b binding; C4b binding; human complement type 1 receptor.
OS Homo sapiens.
FH Key Location/Qualifiers
FT region 1..60
FT /label= SCR-1
FT 61..122
FT /label= SCR-2
FT 451..510
FT /label= SCR-8
FT 511..543
FT /label= SCR-9
FT /note= "TRUNCATED"
FT misc-difference 35
FT /note= "Gly substituted by Glu (SCR-8)"
FT EP-512733-A.
FT 11-NOV-1992.
FT 28-APR-1992; 303826.
FT 03-MAY-1991; US-695514.
FT (UNITW ) UNIV WASHINGTON.
FT Atkinson JP, Hourcade D, Krych M;
FT MPI: 92-375009/46.
FT DR
PT Complement activity regulator protein analogues - useful for
PT treating auto-immune diseases, to suppress transplant rejection,
PT for diagnosis etc.
PT Claim 11; Fig 2 and R11810; 23pp; English.
CC The CDNA clone designated CRI-4 was described in J.Exp.Med.(1988)
CC 168:1255-1270. It encodes the first 8 and a half amino terminal
CC SCRs of CRI. The invention concerns analogues of "regulator of
CC complement activation" proteins or truncated, hybrid or recombinant
CC forms of them. CRI-4 is a preferred truncated form and a number of
CC specified substitution variants of it are claimed. Positions 35 and
CC 37 of SCR-1 and the corresponding positions in SCR-8 have been
CC identified as important in C4b binding. The specification does not
CC contain the CRI-4 sequence; the sequence given here was constructed
CC from the full-length CRI amino acid sequence having GENSEQ
CC accession number R11810 and descriptions in the disclosure.
CC Sequence 543 AA;
SQ

Query Match 100.0%; Score 91; DB 1; Length 543;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SGGKRVFELVGPSTYC 17
Db 158 SGGKRVFELVGPSTYC 174

RESULT 6
R28545
ID R28545 standard; peptide: 543 AA.
AC R28545;
DT 19-MAR-1993 (first entry)
DE CRI-4 (37Y) analogue.
KW short consensus repeat; regulator of complement activation;
  C3b binding; C4b binding; human complement type 1 receptor.
OS Homo sapiens.
FH Key Location/Qualifiers
FT region 1..60
FT /label= SCR-1
FT 61..122
FT /label= SCR-2
FT 451..510
FT /label= SCR-8
FT 511..543
FT /label= SCR-9
FT /note= "TRUNCATED"
FT

```

```

FT misc-difference 37
FT /note= "Ser substituted by Tyr (SCR-8)"
FT EP-512733-A.
FT 11-NOV-1992.
FT 28-APR-1992; 303826.
FT 03-MAY-1991; US-695514.
FT (UNITW ) UNIV WASHINGTON.
FT Atkinson JP, Hourcade D, Krych M;
FT MPI: 92-375009/46.
FT DR
PT Complement activity regulator protein analogues - useful for
PT treating auto-immune diseases, to suppress transplant rejection,
PT for diagnosis etc.
PT Claim 11; Fig 2 and R11810; 23pp; English.
CC The CDNA clone designated CRI-4 was described in J.Exp.Med.(1988)
CC 168:1255-1270. It encodes the first 8 and a half amino terminal
CC SCRs of CRI. The invention concerns analogues of "regulator of
CC complement activation" proteins or truncated, hybrid or recombinant
CC forms of them. CRI-4 is a preferred truncated form and a number of
CC specified substitution variants of it are claimed. Positions 35 and
CC 37 of SCR-1 and the corresponding positions in SCR-8 have been
CC identified as important in C4b binding. The specification does not
CC contain the CRI-4 sequence; the sequence given here was constructed
CC from the full-length CRI amino acid sequence having GENSEQ
CC accession number R11810 and descriptions in the disclosure.
CC Sequence 543 AA;
SQ

Query Match 100.0%; Score 91; DB 1; Length 543;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SGGKRVFELVGPSTYC 17
Db 158 SGGKRVFELVGPSTYC 174

RESULT 7
R28546
ID R28546 standard; peptide: 543 AA.
AC R28546;
DT 19-MAR-1993 (first entry)
DE CRI-4 (44T, 47D, 49L) analogue.
KW short consensus repeat; regulator of complement activation;
  C3b binding; C4b binding; human complement type 1 receptor.
OS Homo sapiens.
FH Key Location/Qualifiers
FT region 1..60
FT /label= SCR-1
FT 61..122
FT /label= SCR-2
FT 451..510
FT /label= SCR-8
FT 511..543
FT /label= SCR-9
FT /note= "TRUNCATED"
FT misc-difference 44
FT /note= "Ile substituted by Thr (SCR-8)"
FT misc-difference 47
FT /note= "Lys substituted by Asp (SCR-8)"
FT misc-difference 49
FT /note= "Ser substituted by Leu (SCR-8)"
FT EP-512733-A.
FT 11-NOV-1992.
FT 28-APR-1992; 303826.
FT 03-MAY-1991; US-695514.
FT (UNITW ) UNIV WASHINGTON.
FT Atkinson JP, Hourcade D, Krych M;
FT MPI: 92-375009/46.
FT DR
PT Complement activity regulator protein analogues - useful for
PT treating auto-immune diseases, to suppress transplant rejection,
PT for diagnosis etc.
PT Claim 11; Fig 2 and R11810; 23pp; English.
CC The CDNA clone designated CRI-4 was described in J.Exp.Med.(1988)

```

CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCRs of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-1 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-8. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENSEQ  
CC accession number R11810 and descriptions in the disclosure.  
SQ Sequence 543 AA;

Query Match 100.0%; Score 91; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 2.3e-08;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGGRRVFLVGEPSIYC 17  
|||||  
DB 158 SGGRRVFLVGEPSIYC 174

RESULT 8  
R28547  
ID R28547 standard; peptide: 543 AA.  
AC R28547;  
DT 19-MAR-1993 (first entry)  
DE CRI-4 (525, 535, 54P) analogue.  
KW Short consensus repeat; regulator of complement activation;  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT region 1..60  
FT /label= SCR-1  
FT region 61..122  
FT /label= SCR-2  
FT region 451..510  
FT /label= SCR-8  
FT region 511..543  
FT /label= SCR-9  
FT /note= "TRUNCATED"  
FT misc-difference 52  
FT /note= "Thr substituted by Ser (SCR-8)"  
FT misc-difference 53  
FT /note= "Gly substituted by Ser (SCR-8)"  
FT misc-difference 54  
FT /note= "Ala substituted by Pro (SCR-8)"  
PN EP-512733-A.  
PD 11-NOV-1992.  
PE 28-APR-1992; 303826.  
PR 03-MAY-1991; US-695514.  
PA (UNIV ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Kirsch M;  
DR WPI: 92-375009/46.  
PT Complement activity regulator protein analogues - useful for  
PT treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11; Fig 2 and R11810; 23pp; English.  
CC The cDNA clone designated CRI-4 was described in J. Exp. Med. (1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCRs of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-1 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-8. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENSEQ  
CC accession number R11810 and descriptions in the disclosure.  
SQ Sequence 543 AA;

Query Match 100.0%; Score 91; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 2.3e-08;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGGRRVFLVGEPSIYC 17  
|||||  
DB 158 SGGRRVFLVGEPSIYC 174

RESULT 9

R28548  
ID R28548 standard; peptide: 543 AA.

AC R28548;

DT 19-MAR-1993 (first entry)

DE CRI-4 (57V, 59K) analogue.

KW Short consensus repeat; regulator of complement activation;  
OS Homo sapiens.

FH Key Location/Qualifiers

FT region 1..60

FT /label= SCR-1

FT region 61..122

FT /label= SCR-2

FT region 451..510

FT /label= SCR-8

FT region 511..543

FT /label= SCR-9

FT /note= "TRUNCATED"

FT misc-difference 57

FT /note= "Arg substituted by Val (SCR-8)"

FT misc-difference 59

FT /note= "Arg substituted by Lys (SCR-8)"

PN EP-512733-A.

PD 11-NOV-1992.

PE 28-APR-1992; 303826.

PR 03-MAY-1991; US-695514.

PA (UNIV ) UNIV WASHINGTON.

PI Atkinson JP, Hourcade D, Kirsch M;

DR WPI: 92-375009/46.

PT Complement activity regulator protein analogues - useful for  
PT treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.

PS Claim 11; Fig 2 and R11810; 23pp; English.

CC The cDNA clone designated CRI-4 was described in J. Exp. Med. (1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCRs of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-1 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-8. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENSEQ  
CC accession number R11810 and descriptions in the disclosure.  
SQ Sequence 543 AA;

Query Match 100.0%; Score 91; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 2.3e-08;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGGRRVFLVGEPSIYC 17  
|||||  
DB 158 SGGRRVFLVGEPSIYC 174

RESULT 10

R28549  
ID R28549 standard; peptide: 543 AA.

AC R28549;

DT 19-MAR-1993 (first entry)



DE CRI-4 (64K, 65T) analogue.  
KW short consensus repeat; regulator of complement activation;  
KW C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT region 1..60  
FT /label= SCR-1  
FT region 61..122  
FT /label= SCR-2  
FT region 451..510  
FT /label= SCR-8  
FT region 511..543  
FT /label= SCR-9  
FT misc\_difference 64  
FT /note= "TRUNCATED"  
FT misc\_difference 65  
FT /note= "Arg substituted by Lys (SCR-9)"  
FT /note= "Asn substituted by Thr (SCR-9)"  
PN EP-512733-A.  
PD 11-NOV-1992.  
PF 28-APR-1992; 303826.  
PR 03-MAY-1991; US-695514.  
PA (UNITV ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Krych M;  
PI MPI; 92-375009/46.  
PT Complement activity regulator protein analogues - useful for  
PT treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11: Fig 2 and R1810; 23pp: English.  
CC The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCRs of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-2 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-9. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQ  
CC accession number R1810 and descriptions in the disclosure.  
SQ Sequence 543 AA:  
  
Query Match 100.0%; Score 91; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 2.3e-08;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SGGKRVFELVGPSTYC 17  
DB 158 SGGKRVFELVGPSTYC 174  
  
RESULT 11  
ID R28550 standard; peptide: 543 AA.  
AC R28550:  
DT 19-MAR-1993 (first entry)  
DE CRI-4 (64K) analogue.  
KW short consensus repeat; regulator of complement activation;  
KW C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT region 1..60  
FT /label= SCR-1  
FT region 61..122  
FT /label= SCR-2  
FT region 451..510  
FT /label= SCR-8  
FT region 511..543  
FT /label= SCR-9  
FT misc\_difference 64  
FT /note= "TRUNCATED"

FT EP-512733-A. /note= "Arg substituted by Lys (SCR-9)"  
PN 11-NOV-1992.  
PD 28-APR-1992; 303826.  
PF 03-MAY-1991; US-695514.  
PR (UNITV ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Krych M;  
PI MPI; 92-375009/46.  
PT Complement activity regulator protein analogues - useful for  
PT treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11: Fig 2 and R1810; 23pp: English.  
CC The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCRs of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-2 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-9. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQ  
CC accession number R1810 and descriptions in the disclosure.  
SQ Sequence 543 AA:  
  
Query Match 100.0%; Score 91; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 2.3e-08;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SGGKRVFELVGPSTYC 17  
DB 158 SGGKRVFELVGPSTYC 174  
  
RESULT 12  
ID R28551 standard; peptide: 543 AA.  
AC R28551:  
DT 19-MAR-1993 (first entry)  
DE CRI-4 (65T) analogue.  
KW short consensus repeat; regulator of complement activation;  
KW C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT region 1..60  
FT /label= SCR-1  
FT region 61..122  
FT /label= SCR-2  
FT region 451..510  
FT /label= SCR-8  
FT region 511..543  
FT /label= SCR-9  
FT misc\_difference 64  
FT /note= "TRUNCATED"  
FT misc\_difference 65  
FT /note= "Arg substituted by Lys (SCR-9)"  
FT /note= "Asn substituted by Thr (SCR-9)"  
PN EP-512733-A.  
PD 11-NOV-1992.  
PF 28-APR-1992; 303826.  
PR 03-MAY-1991; US-695514.  
PA (UNITV ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Krych M;  
PI MPI; 92-375009/46.  
PT Complement activity regulator protein analogues - useful for  
PT treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11: Fig 2 and R1810; 23pp: English.  
CC The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCRs of CRI. The invention concerns analogues of "regulator of

CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-2 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-9. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQ  
CC accession number R11810 and descriptions in the disclosure.  
SO Sequence 543 AA;

Query Match 100.0%; Score 91; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 2.3e-08;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SGGKRVFELVGEPSIYC 17  
Db 158 SGGKRVFELVGEPSIYC 174

## RESULT 13

R28552  
ID R28552 standard; peptide; 543 AA.  
AC R28552;  
DE 19-MAR-1993 (first entry)  
DE CRI-4 (78T, 79D) analogue.  
KW short consensus repeat; regulator of complement activation;  
KW C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT region 1..60  
FT /label= SCR-1  
FT region 61..122  
FT /label= SCR-2  
FT region 451..510  
FT /label= SCR-8  
FT region 511..543  
FT /label= SCR-9  
FT /note="TRUNCATED"  
FT misc\_difference 78  
FT /note="Lys substituted by Thr (SCR-9)"  
FT misc\_difference 79  
FT /note="Gly substituted by Asp (SCR-9)"  
PN EP-512733-A.  
PD 11-NOV-1992.  
PE 28-APR-1992; 303826.  
PR 03-MAY-1991; US-695514.  
PA (UNIV ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Kirsch M;  
DR WPI: 92-375009/46.  
PT Complement activity regulator protein analogues - useful for  
PT treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11; Fig 2 and R11810; 23pp; English.  
CC The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCRs of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-2 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-9. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQ  
CC accession number R11810 and descriptions in the disclosure.  
SO Sequence 543 AA;

Query Match 100.0%; Score 91; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 2.3e-08;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SGGKRVFELVGEPSIYC 17  
Db 158 SGGKRVFELVGEPSIYC 174

## RESULT 14

R28553  
ID R28553 standard; peptide; 543 AA.  
AC R28553;  
DE 19-MAR-1993 (first entry)  
DE CRI-4 (85R, 87N) analogue.  
KW short consensus repeat; regulator of complement activation;  
KW C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT region 1..60  
FT /label= SCR-1  
FT region 61..122  
FT /label= SCR-2  
FT region 451..510  
FT /label= SCR-8  
FT region 511..543  
FT /label= SCR-9  
FT /note="TRUNCATED"  
FT misc\_difference 85  
FT /note="Gln substituted by Arg (SCR-9)"  
FT misc\_difference 87  
FT /note="Lys substituted by Asn (SCR-9)"  
PN EP-512733-A.  
PD 11-NOV-1992.  
PE 28-APR-1992; 303826.  
PR 03-MAY-1991; US-695514.  
PA (UNIV ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Kirsch M;  
DR WPI: 92-375009/46.  
PT Complement activity regulator protein analogues - useful for  
PT treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11; Fig 2 and R11810; 23pp; English.  
CC The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCRs of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-2 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-9. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQ  
CC accession number R11810 and descriptions in the disclosure.  
SO Sequence 543 AA;

Query Match 100.0%; Score 91; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 2.3e-08;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SGGKRVFELVGEPSIYC 17  
Db 158 SGGKRVFELVGEPSIYC 174

## RESULT 15

R28554  
ID R28554 standard; peptide; 543 AA.  
AC R28554;  
DE 19-MAR-1993 (first entry)  
DE CRI-4 (92T, 94H) analogue.  
KW short consensus repeat; regulator of complement activation;  
KW C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.





GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 06:07:56 ; Search time 39.54 Seconds  
(without alignments)  
4.243 Million cell updates/sec

Title: US-09-142-043-5

Perfect score: 91

Sequence: 1 SGGRRVFEIIVGPSTYC 17

Scoring table: BLOSUM62

Searched: 106577 seqs, 9868381 residues

Database : Issued\_Patents\_AA:\*

1: /cgn2\_6/prodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/PCITUS9.COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/Backfilltest1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	91	100.0	197	2	US-08-356-361-27	Sequence 27, Appl
2	91	100.0	76	2	US-08-356-361-28	Sequence 28, Appl
3	91	100.0	254	2	US-08-356-361-29	Sequence 29, Appl
4	91	100.0	254	2	US-08-356-361-30	Sequence 30, Appl
5	91	100.0	133	2	US-08-356-361-31	Sequence 31, Appl
6	91	100.0	197	2	US-08-769-967A-27	Sequence 27, Appl
7	91	100.0	76	2	US-08-769-967A-28	Sequence 28, Appl
8	91	100.0	254	2	US-08-769-967A-29	Sequence 29, Appl
9	91	100.0	254	2	US-08-769-967A-30	Sequence 30, Appl
10	91	100.0	133	2	US-08-769-967A-31	Sequence 31, Appl
11	55	60.4	310	4	5256642-10	Patent No. 5256642
12	55	60.4	310	4	5472939-10	Patent No. 5472939
13	55	54.9	260	4	5256642-5	Patent No. 5256642
14	50	54.9	250	4	5256642-6	Patent No. 5256642
15	50	54.9	260	4	5472939-5	Patent No. 5472939
16	50	54.9	250	4	5472939-6	Patent No. 5472939
17	49	53.8	10	4	5256642-18	Patent No. 5256642
18	49	53.8	10	4	5472939-18	Patent No. 5472939
19	45	49.5	254	1	US-08-310-416A-13	Sequence 13, Appl
20	45	49.5	324	1	US-08-310-416A-14	Sequence 14, Appl
21	45	49.5	293	1	US-08-310-416A-16	Sequence 16, Appl
22	45	49.5	169	1	US-08-310-416A-18	Sequence 18, Appl
23	45	49.5	377	2	US-08-528-057-2	Sequence 2, Appl
24	45	49.5	370	2	US-08-528-057-42	Sequence 42, Appl
25	45	49.5	373	2	US-08-528-057-44	Sequence 44, Appl
26	45	49.5	324	2	US-08-528-057-46	Sequence 46, Appl
27	45	49.5	254	2	US-08-888-171-13	Sequence 13, Appl
28	45	49.5	324	2	US-08-888-171-14	Sequence 14, Appl
29	45	49.5	293	2	US-08-888-171-16	Sequence 16, Appl
30	45	49.5	169	2	US-08-888-171-18	Sequence 18, Appl
31	45	49.5	254	2	US-08-435-149-1	Sequence 1, Appl
32	45	49.5	323	2	US-08-435-149-2	Sequence 2, Appl
33	45	49.5	577	2	US-08-435-149-3	Sequence 3, Appl
34	45	49.5	29	2	US-08-637-759B-353	Sequence 353, App
35	37	40.7	752	2	US-08-896-590A-2	Sequence 2, Appl
36	37	40.7	752	2	US-08-896-590A-4	Sequence 4, Appl
37	37	40.7	340	4	5256642-2	Patent No. 5256642
38	37	40.7	340	4	5472939-2	Patent No. 5472939
39	36	39.6	293	1	US-08-446-925-5	Sequence 5, Appl

40	36	39.6	204	1	US-08-446-925-7	Sequence 7, Appl
41	36	39.6	591	2	US-08-889-402-1	Sequence 1, Appl
42	36	39.6	605	2	US-08-889-402-2	Sequence 2, Appl
43	36	39.6	293	3	PCR-US96-10521-31	Sequence 31, Appl
44	35	38.5	951	1	US-08-162-809-2	Sequence 2, Appl
45	35	38.5	984	2	US-08-673-789-6	Sequence 6, Appl

## ALIGNMENTS

```
RESULT 1
US-08-356-361-27
; Sequence 27, Application US/08356361
; Patent No. 5833989
; GENERAL INFORMATION:
; APPLICANT: Smith, Richard A.G.
; APPLICANT: Dodd, Ian
; APPLICANT: Freeman Mary A.
; APPLICANT: Mossakowska, Danuta E.I.
; TITLE OF INVENTION: No. 5833989el Compounds
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham-Corporate Intellectual Property
; STREET: P.O. Box 1539
; CITY: King of Prussia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/356,361
; FILING DATE: 03-Jul-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jervis, Herbert H.
; REGISTRATION NUMBER: 31,171
; REFERENCE/DOCKET NUMBER: P30423
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (610) 270-5019
; TELEFAX: (610) 270-5090
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 197 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: N-terminal
; US-08-356-361-27

Query Match 100.0%; Score 91; DB 2; Length 197;
Best local similarity 100.0%; Pred. No. 3.7e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGGRRVFEIIVGPSTYC 17
Db 159 SGGRRVFEIIVGPSTYC 175

RESULT 2
US-08-356-361-28
; Sequence 28, Application US/08356361
; Patent No. 5833989
; GENERAL INFORMATION:
; APPLICANT: Smith, Richard A.G.
; APPLICANT: Dodd, Ian
; APPLICANT: Freeman Mary A.
; APPLICANT: Mossakowska, Danuta E.I.
```

TITLE OF INVENTION: No. 5833989e1 Compounds  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham-Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,361  
FILING DATE: 03-Jul-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jervis, Herbert H.  
REGISTRATION NUMBER: 31,171  
REFERENCE/DOCKET NUMBER: P30423  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5019  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 76 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-356-361-28

Query Match 100.0%; Score 91; DB 2; Length 76;  
Best Local Similarity 100.0%; Pred. No. 1.2e-09;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGGKRVFELVGEPSIYC 17  
|||||  
DB 38 SGGKRVFELVGEPSIYC 54

RESULT 3  
US-08-356-361-29  
Sequence 29, Application US/08356361  
Patent No. 5833989  
GENERAL INFORMATION:  
APPLICANT: Smith, Richard A. G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman Mary A.  
APPLICANT: Mossakowska, Danuta E. I.  
TITLE OF INVENTION: No. 5833989e1 Compounds  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham-Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,361  
FILING DATE: 03-Jul-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jervis, Herbert H.

REGISTRATION NUMBER: 31,171  
REFERENCE/DOCKET NUMBER: P30423  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5019  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-356-361-29

Query Match 100.0%; Score 91; DB 2; Length 254;  
Best Local Similarity 100.0%; Pred. No. 4.9e-09;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGGKRVFELVGEPSIYC 17  
|||||  
DB 159 SGGKRVFELVGEPSIYC 175

RESULT 4  
US-08-356-361-30  
Sequence 30, Application US/08356361  
Patent No. 5833989  
GENERAL INFORMATION:  
APPLICANT: Smith, Richard A. G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman Mary A.  
APPLICANT: Mossakowska, Danuta E. I.  
TITLE OF INVENTION: No. 5833989e1 Compounds  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham-Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,361  
FILING DATE: 03-Jul-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jervis, Herbert H.  
REGISTRATION NUMBER: 31,171  
REFERENCE/DOCKET NUMBER: P30423  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5019  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-356-361-30

Query Match 100.0%; Score 91; DB 2; Length 254;  
Best Local Similarity 100.0%; Pred. No. 4.9e-09;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGGKRVFELVGEPSIYC 17

DB 159 SSGRKVFELVGEPSIYC 175

## RESULT 5

US-08-356-361-31  
Sequence 31, Application US/08356361  
Patent No. 5833989  
GENERAL INFORMATION:  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman Mary A.  
APPLICANT: Mossakowska, Danuta E.I.  
TITLE OF INVENTION: NO. 5833989el Compounds  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham-Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,361  
FILING DATE: 03-Jul-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jarvis, Herbert H.  
REGISTRATION NUMBER: 31,171  
REFERENCE/DOCKET NUMBER: P30423  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5019  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 133 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-356-361-31

Query Match 100.0%; Score 91; DB 2; Length 133;  
Best Local Similarity 100.0%; Pred. No. 2.3e-09;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSGRKVFELVGEPSIYC 17  
DB 38 SSGRKVFELVGEPSIYC 54

## RESULT 6

US-08-769-967A-27  
Sequence 27, Application US/0876967A  
Patent No. 5859223  
GENERAL INFORMATION:  
APPLICANT: Mossakowska, Danuta E.I.  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman, Anne Mary  
TITLE OF INVENTION: Soluble CRI Derivatives  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania

COUNTRY: USA  
ZIP: 19406

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,967A  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,569  
FILING DATE: 15-May-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: P30423C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5364  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 197 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-769-967A-27

Query Match 100.0%; Score 91; DB 2; Length 197;  
Best Local Similarity 100.0%; Pred. No. 3.7e-09;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSGRKVFELVGEPSIYC 17  
DB 159 SSGRKVFELVGEPSIYC 175

## RESULT 7

US-08-769-967A-28  
Sequence 28, Application US/0876967A  
Patent No. 5859223  
GENERAL INFORMATION:  
APPLICANT: Mossakowska, Danuta E.I.  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman, Anne Mary  
TITLE OF INVENTION: Soluble CRI Derivatives  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,967A  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,569  
FILING DATE: 15-May-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954

REFERENCE/DOCKET NUMBER: P30423C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5364  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 76 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-769-967A-28

Query Match 100.0%; Score 91; DB 2; Length 76;  
Best Local Similarity 100.0%; Pred. No. 1.2e-09;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SGRKRVFELVGEPSIYC 17  
Db 38 SGRKRVFELVGEPSIYC 54

RESULT 8  
US-08-769-967A-29  
Sequence 29, Application US/08769967A  
Patent No. 5859223  
GENERAL INFORMATION:  
APPLICANT: Mossakowska, Danuta E.I.  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman, Anne Mary  
TITLE OF INVENTION: Soluble CRI Derivatives  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,967A  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,569  
FILING DATE: 15-May-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: P30423C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5364  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-769-967A-29

Query Match 100.0%; Score 91; DB 2; Length 254;  
Best Local Similarity 100.0%; Pred. No. 4.9e-09;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SGRKRVFELVGEPSIYC 17  
Db 159 SGRKRVFELVGEPSIYC 175

RESULT 9  
US-08-769-967A-30  
Sequence 30, Application US/08769967A  
Patent No. 5859223  
GENERAL INFORMATION:  
APPLICANT: Mossakowska, Danuta E.I.  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman, Anne Mary  
TITLE OF INVENTION: Soluble CRI Derivatives  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,967A  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,569  
FILING DATE: 15-May-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: P30423C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5364  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-769-967A-30

Query Match 100.0%; Score 91; DB 2; Length 254;  
Best Local Similarity 100.0%; Pred. No. 4.9e-09;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SGRKRVFELVGEPSIYC 17  
Db 159 SGRKRVFELVGEPSIYC 175

RESULT 10  
US-08-769-967A-31  
Sequence 31, Application US/08769967A  
Patent No. 5859223  
GENERAL INFORMATION:  
APPLICANT: Mossakowska, Danuta E.I.  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman, Anne Mary  
TITLE OF INVENTION: Soluble CRI Derivatives  
NUMBER OF SEQUENCES: 33



CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmitKline Beecham Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King Of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,967A  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,569  
FILING DATE: 15-May-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: P30423C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5364  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 133 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-769-967A-31

Query Match 100.0%; Score 91; DB 2; Length 133;  
Best Local Similarity 100.0%; Pred. No. 2.3e-09;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGGKRVFELVGEPSIYC 17  
|||||  
Db 38 SGGKRVFELVGEPSIYC 54

RESULT 11  
5256642-10  
PATENT NO. 5256642  
APPLICANT: FEARON, DOUGLAS T.; KLICKSTEIN, LLOYD B.; WONG,  
WINNIE W.; CARSON, GERALD R.; CONCINO, MICHAEL F.; IP, STEPHEN  
H.; MAKRIDES, SAVVAS; MARSH, HENRY C. JR.  
TITLE OF INVENTION: COMPOSITIONS OF SOLUBLE COMPLEMENT  
RECEPTOR 1 (CRI) AND A THROMBOLYTIC AGENT, AND THE METHODS OF  
USE THEREOF  
NUMBER OF SEQUENCES: 30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/588,128  
FILING DATE: 24-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 412,745  
FILING DATE: 26-SEP-1989  
APPLICATION NUMBER: 332,865  
FILING DATE: 03-APR-1989  
APPLICATION NUMBER: 176,532  
FILING DATE: 01-APR-1988  
SEQ ID NO: 10;  
LENGTH: 1847  
5256642-10

Query Match 60.4%; Score 55; DB 4; Length 310;  
Best Local Similarity 100.0%; Pred. No. 0.019;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 ELVGEPSIYC 17  
|||||  
Db 111 ELVGEPSIYC 120

RESULT 12  
5472939-10  
PATENT NO. 5472939  
APPLICANT: FEARON, DOUGLAS T.; KLICKSTEIN, LLOYD B.; WONG,  
WINNIE W.; CARSON, GERALD R.; CONCINO, MICHAEL F.; IP, STEPHEN  
H.; MAKRIDES, SAVVAS; MARSH, HENRY C. JR.  
TITLE OF INVENTION: METHOD OF TREATING COMPLEMENT  
MEDIATED DISORDERS  
NUMBER OF SEQUENCES: 30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/138,825  
FILING DATE: 19-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 588,128  
FILING DATE: 24-SEP-1990  
APPLICATION NUMBER: 412,745  
FILING DATE: 26-SEP-1989  
APPLICATION NUMBER: 332,865  
FILING DATE: 03-APR-1989  
APPLICATION NUMBER: 176,532  
FILING DATE: 01-APR-1988  
SEQ ID NO: 10;  
LENGTH: 2006  
5472939-10

Query Match 60.4%; Score 55; DB 4; Length 310;  
Best Local Similarity 100.0%; Pred. No. 0.019;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 ELVGEPSIYC 17  
|||||  
Db 111 ELVGEPSIYC 120

RESULT 13  
5256642-5  
PATENT NO. 5256642  
APPLICANT: FEARON, DOUGLAS T.; KLICKSTEIN, LLOYD B.; WONG,  
WINNIE W.; CARSON, GERALD R.; CONCINO, MICHAEL F.; IP, STEPHEN  
H.; MAKRIDES, SAVVAS; MARSH, HENRY C. JR.  
TITLE OF INVENTION: COMPOSITIONS OF SOLUBLE COMPLEMENT  
RECEPTOR 1 (CRI) AND A THROMBOLYTIC AGENT, AND THE METHODS OF  
USE THEREOF  
NUMBER OF SEQUENCES: 30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/588,128  
FILING DATE: 24-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 412,745  
FILING DATE: 26-SEP-1989  
APPLICATION NUMBER: 332,865  
FILING DATE: 03-APR-1989  
APPLICATION NUMBER: 176,532  
FILING DATE: 01-APR-1988  
SEQ ID NO: 5;  
LENGTH: 1537  
5256642-5

Query Match 54.9%; Score 50; DB 4; Length 260;  
Best Local Similarity 100.0%; Pred. No. 0.12;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 KVFELVGEPS 14  
|||||  
Db 101 KVFELVGEPS 110

Job time: 298 sec

```
RESULT 14
5256642-6
; Patent No. 5256642
; APPLICANT: FEARON, DOUGLAS T.; KICKSTEIN, LLOYD B.; WONG,
; WINNIE W.; CARSON, GERALD R.; CONCINO, MICHAEL F.; IP, STEPHEN
; H.; MAKRIDES, SAVVAS; MARSH, HENRY C. JR.
; TITLE OF INVENTION: COMPOSITIONS OF SOLUBLE COMPLEMENT
; RECEPTOR 1 (CRL) AND A THROMBOLYTIC AGENT, AND THE METHODS OF
; USE THEREOF
; NUMBER OF SEQUENCES: 30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/588,128
; FILING DATE: 24-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 412,745
; FILING DATE: 26-SEP-1989
; APPLICATION NUMBER: 332,865
; FILING DATE: 03-APR-1989
; APPLICATION NUMBER: 176,532
; FILING DATE: 01-APR-1988
; SEQ ID NO: 6:
; LENGTH: 1466
5256642-6

Query Match 54.9%; Score 50; DB 4; Length 250;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 KVFELVGEPS 14
|||||
DB 101 KVFELVGEPS 110

RESULT 15
5472939-5
; Patent No. 5472939
; APPLICANT: FEARON, DOUGLAS T.; KICKSTEIN, LLOYD B.; WONG,
; WINNIE W.; CARSON, GERALD R.; CONCINO, MICHAEL F.; IP, STEPHEN
; H.; MAKRIDES, SAVVAS; MARSH, HENRY C. JR.
; TITLE OF INVENTION: METHOD OF TREATING COMPLEMENT
; MEDIATED DISORDERS
; NUMBER OF SEQUENCES: 30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/138,825
; FILING DATE: 19-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 588,128
; FILING DATE: 24-SEP-1990
; APPLICATION NUMBER: 412,745
; FILING DATE: 26-SEP-1989
; APPLICATION NUMBER: 332,865
; FILING DATE: 03-APR-1989
; APPLICATION NUMBER: 176,532
; FILING DATE: 01-APR-1988
; SEQ ID NO: 5:
; LENGTH: 1537
5472939-5

Query Match 54.9%; Score 50; DB 4; Length 260;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 KVFELVGEPS 14
|||||
DB 101 KVFELVGEPS 110
```

Search completed: July 18, 1999, 06:07:56

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 06:07:04 ; Search time 44.47 Seconds  
(without alignments)  
14.320 Million cell updates/sec

Title: US-09-142-043-5

Perfect score: 91

Sequence: 1 SGRKRVFELVGEPSIYC 17

Scoring table: BLOSUM62

Searched: 116738 seqs, 37460341 residues

Database : PIR\_58:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	91	100.0	2014	2	I36936	complement recepto
2	91	100.0	661	2	I36937	complement recepto
3	91	100.0	2489	2	I73012	complement C3b/C4b
4	91	100.0	2039	2	A28507	complement C3b/C4b
5	80	87.9	482	2	A34924	complement C3b/C4b
6	65	71.4	497	2	JC2054	complement regulat
7	62	68.1	433	2	A30550	complement C3b/C4b
8	62	68.1	440	2	A43519	complement recepto
9	56	61.5	676	2	A45900	complement C3b rec
10	45	49.5	381	1	B26359	decay-accelerating
11	45	49.5	440	2	A26359	decay-accelerating
12	45	49.5	369	2	I57988	membrane cofactor
13	45	49.5	384	2	S01896	membrane cofactor
14	45	49.5	377	2	I54479	membrane cofactor
15	45	49.5	349	2	G02913	sperm CD46 - human
16	44	48.4	611	2	F70325	conserved hypotet
17	43.5	47.8	156	2	A48865	8-oxo-7,8-dihydrog
18	43	47.3	274	2	S22308	type II site-speci
19	43	47.3	369	2	JC5138	membrane cofactor
20	43	47.3	362	2	JC5194	membrane cofactor
21	42	46.2	165	2	I41043	dihydrofolate redu
22	42	46.2	165	2	S60665	dihydrofolate redu
23	42	46.2	159	2	S32014	dihydrofolate redu
24	42	46.2	340	2	I56234	dihydrofolate redu
25	41	45.1	168	1	RDBSD	decay-accelerating
26	41	45.1	160	2	C64101	dihydrofolate redu
27	41	45.1	160	2	S52338	dihydrofolate redu
28	41	45.1	160	2	S52338	dihydrofolate redu
29	41	45.1	666	2	D69103	DNA replication in
30	41	45.1	724	2	C71274	hypothetical prote
31	40	44.0	1025	2	A43526	probable thiolosulf
32	40	44.0	277	2	G70809	prochlorophyllid
33	40	44.0	313	2	S08406	precocin-3 methyl
34	39.5	43.4	259	2	A64497	polyketide synthas
35	39.5	43.4	341	1	JC5855	DNA-directed DNA p
36	39	42.9	1061	1	D1912	conserved hypotet
37	39	42.9	225	2	E69256	hypothetical prote
38	39	42.9	153	2	E69403	hypothetical prote
39	39	42.9	4436	2	E71086	hypothetical prote

40	38.5	42.3	291	2	S17741	fixc protein - sho
41	38.5	42.3	671	2	A55164	anaphase control p
42	38.5	42.3	321	2	S62015	SOI protein - yea
43	38	41.8	162	1	RDEBDT	dihydrofolate redu
44	38	41.8	205	1	PQ0032	collin E9 (EC 3.1
45	38	41.8	477	1	DWDXAF	2-hydroxyglutaryl-

ALIGNMENTS

RESULT 1  
I36936  
complement receptor I - chimpanzee (fragment)  
C:Species: Pan troglodytes (chimpanzee)  
C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 07-Aug-1998  
C:Accession: I36936  
R:Birmingham, D.J.; Shen, X.P.; Hourcade, D.; Nickells, M.W.; Atkinson, J.P.  
J. Immunol. 153, 691-700, 1994  
A:Title: Primary sequence of an alternatively spliced form of CRI. Candidate for the  
A:Reference number: I36935; MUID:94292799  
A:Accession: I36936  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-2014 <RES>  
A:Cross-references: GB:I24920; NID:9551564; PID:9557725  
C:Superfamily: complement factor H repeat homology  
F:18-74/Domain: complement factor H repeat homology <FH01>  
F:79-136/Domain: complement factor H repeat homology <FH02>  
F:141-207/Domain: complement factor H repeat homology <FH03>  
F:333-391/Domain: complement factor H repeat homology <FH12>  
F:1041-1107/Domain: complement factor H repeat homology <FH26>  
F:1745-1815/Domain: complement factor H repeat homology <FH27>

Query Match 100.0%; Score 91; DB 2; Length 2014;  
Best Local Similarity 100.0%; Pred. No. 2.2e-07;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGRKRVFELVGEPSIYC 17  
DB 174 SGRKRVFELVGEPSIYC 190

RESULT 2  
I36937  
Complement receptor I - chimpanzee (fragment)  
C:Species: Pan troglodytes (chimpanzee)  
C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 07-Aug-1998  
C:Accession: I36937  
R:Birmingham, D.J.; Shen, X.P.; Hourcade, D.; Nickells, M.W.; Atkinson, J.P.  
J. Immunol. 153, 691-700, 1994  
A:Title: Primary sequence of an alternatively spliced form of CRI. Candidate for the  
A:Reference number: I36935; MUID:94292799  
A:Accession: I36937  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-661 <RES>  
A:Cross-references: GB:I24921; NID:9557726; PID:9557727  
C:Superfamily: complement factor H repeat homology  
F:18-74/Domain: complement factor H repeat homology <FH01>  
F:79-136/Domain: complement factor H repeat homology <FH02>  
F:141-207/Domain: complement factor H repeat homology <FH03>  
F:213-268/Domain: complement factor H repeat homology <FH04>  
F:333-391/Domain: complement factor H repeat homology <FH22>  
F:396-462/Domain: complement factor H repeat homology <FH27>

Query Match 100.0%; Score 91; DB 2; Length 661;  
Best Local Similarity 100.0%; Pred. No. 6.9e-08;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGRKRVFELVGEPSIYC 17

Db 174 SGRKRVFELVGPSTYC 190  
|||||  
RESULT 3  
173012  
Complement C3b/C4b receptor (allotype S) precursor - human  
N:Alternate names: complement receptor type 1 (CR1); surface glycoprotein CD35  
C:Species: Homo sapiens (man)  
C:Date: 02-Jul-1996 #sequence,revision 02-Jul-1996 #text\_change 24-Sep-1998  
C:Accession: 173012; A47602; S03291  
R:Vik, D.P.; Wong, W.W.  
J:Immunol. 151, 6214-6224, 1993  
A:Title: Structure of the gene for the F allele of complement receptor type 1 and sequen  
A:Reference number: 156203; MUID:94065175  
A:Accession: 173012  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2489 <RES>  
A:Cross-references: GB:L17418; NID:g306678; PID:9451303  
J:Wong, W.W.; Cahill, J.M.; Rosen, M.D.; Kennedy, C.A.; Bonaccio, E.T.; Morris, M.J.; Wi  
J:Exp. Med. 169, 847-863, 1989  
A:Title: Structure of the human CR1 gene. Molecular basis of the structural and quantita  
A:Reference number: A47602  
A:Accession: A47602  
A:Molecule type: DNA  
A:Residues: 1-41 <WON>  
A:Cross-references: GB:X14893  
R:Hourcade, D.; Miesner, D.R.; Atkinson, J.P.; Holers, V.M.  
J:Exp. Med. 168, 1255-1270, 1988  
A:Title: Identification of an alternative polyadenylation site in the human C3b/C4b rece  
type 1.  
A:Reference number: S03291; MUID:89010527  
A:Accession: S03291  
A:Molecule type: mRNA  
A:Residues: 26-584 <HOU>  
A:Cross-references: EMBL:X14362; NID:g30197; PID:g736240  
A:Experimental source: clone CR1-4  
C:Genetics:  
A:Gene: GDB:CR1; CD35  
A:Cross-references: GDB:119800; OMIM:120620  
A:Map position: 1q32-1q32  
A:Introns: 41/1; 101/1; 134/2; 163/1; 296/1; 355/1; 389/2; 418/1; 491/1; 551/1; 584/2; 6  
1/1484/2; 1513/1; 1646/1; 1705/1; 1739/2; 1768/1; 1844/1; 1904/1; 1937/2; 1966/1; 2099/1  
C:Superfamily: complement factor H repeat homology  
C:Keywords: duplication; glycoprotein; transmembrane protein  
F:1-41/Domain: signal sequence #status predicted <SIG>  
F:42-584/Product: complement C3b/C4b receptor, secreted #status predicted <MAT>  
F:43-99/Domain: complement factor H repeat homology <FH01>  
F:104-161/Domain: complement factor H repeat homology <FH02>  
F:164-233/Domain: complement factor H repeat homology <FH03>  
F:238-293/Domain: complement factor H repeat homology <FH04>  
F:297-353/Domain: complement factor H repeat homology <FH05>  
F:358-416/Domain: complement factor H repeat homology <FH06>  
F:421-487/Domain: complement factor H repeat homology <FH07>  
F:554-611/Domain: complement factor H repeat homology <FH08>  
F:616-682/Domain: complement factor H repeat homology <FH09>  
F:808-866/Domain: complement factor H repeat homology <FH09>  
F:1004-1061/Domain: complement factor H repeat homology <FH09>  
F:1066-1132/Domain: complement factor H repeat homology <FH10>  
F:1138-1193/Domain: complement factor H repeat homology <FH11>  
F:1197-1253/Domain: complement factor H repeat homology <FH12>  
F:1258-1316/Domain: complement factor H repeat homology <FH13>  
F:1321-1387/Domain: complement factor H repeat homology <FH14>  
F:1393-1449/Domain: complement factor H repeat homology <FH15>  
F:1454-1511/Domain: complement factor H repeat homology <FH16>  
F:1516-1582/Domain: complement factor H repeat homology <FH17>  
F:1588-1643/Domain: complement factor H repeat homology <FH18>  
F:1647-1703/Domain: complement factor H repeat homology <FH19>  
F:1708-1766/Domain: complement factor H repeat homology <FH20>  
F:1771-1837/Domain: complement factor H repeat homology <FH21>  
F:1846-1902/Domain: complement factor H repeat homology <FH22>  
F:1907-1964/Domain: complement factor H repeat homology <FH23>

F:1969-2035/Domain: complement factor H repeat homology <FHXC>  
F:2100-2156/Domain: complement factor H repeat homology <FH24>  
F:2161-2219/Domain: complement factor H repeat homology <FHXD>  
Query Match 100.0%; Score 91; DB 2; Length 2489,  
Best Local Similarity 100.0%; Pred. No. 2,8e-07;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 SGRKRVFELVGPSTYC 17  
Db 199 SGRKRVFELVGPSTYC 215  
RESULT 4  
A28507  
Complement C3b/C4b receptor precursor, membrane-bound (allotype F) - human  
N:Alternate names: complement receptor type 1 (CR1); surface glycoprotein CD35  
C:Species: Homo sapiens (man)  
C:Date: 19-Nov-1988 #sequence,revision 06-Sep-1996 #text\_change 10-Sep-1997  
C:Accession: S03843; A28507; 156203; A24748; B24748; C24748  
R:Klickstein, L.B.; Barrow, T.J.; Miletic, V.; Rabson, L.D.; Smith, J.A.; Fearon, D.T  
J:Exp. Med. 168, 1699-1717, 1988  
A:Title: Identification of distinct C3b and C4b recognition sites in the human C3b/C4  
A:Reference number: S03843; MUID:89035992  
A:Accession: S03843  
A:Status: translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-2039 <KLI>  
A:Cross-references: EMBL:Y00816; NID:g30185; PID:g30186  
R:Klickstein, L.B.; Wong, W.W.; Smith, J.A.; Weis, J.H.; Wilson, J.G.; Fearon, D.T.  
J:Exp. Med. 165, 1095-1112, 1987  
A:Title: Human C3b/C4b receptor (CR1). Demonstration of long homologous repeating dom  
A:Reference number: A28507; MUID:87168191  
A:Accession: A28507  
A:Molecule type: mRNA  
A:Residues: 503-771, 'FV', 774-2039 <KLI>  
R:Vik, D.P.; Wong, W.W.  
J:Immunol. 151, 6214-6224, 1993  
A:Title: Structure of the gene for the F allele of complement receptor type 1 and seq  
A:Reference number: 156203; MUID:94065175  
A:Accession: 156203  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-683, 'X', 685-1021, 'X', 1023-1614, 'V', 1616-1826, 'R', 1828-1849, 'D', 1851-187  
A:Cross-references: GB:L17418; NID:g306678; PID:g306680  
R:Wong, W.W.; Klickstein, L.B.; Smith, J.A.; Weis, J.H.; Fearon, D.T.  
Proc. Natl. Acad. Sci. U.S.A. 82, 7711-7715, 1985  
A:Title: Identification of a partial cDNA clone for the human receptor for complement  
A:Reference number: A94073; MUID:86067975  
A:Accession: A24748  
A:Molecule type: mRNA  
A:Residues: 311-333; 729-745; 831-845 <WON>  
C:Genetics:  
A:Gene: GDB:CR1; CD35  
A:Cross-references: GDB:119800; OMIM:120620  
A:Map position: 1q32-1q32  
A:Introns: 41/1; 101/1; 134/2; 163/1; 296/1; 355/1; 389/2; 418/1; 491/1; 551/1; 584/2  
1/1487/2; 1516/1; 1649/1; 1708/1; 1742/2; 1771/1; 1841/1; 1906/1; 1966/1; 1976/1; 200  
C:Superfamily: complement factor H repeat homology  
C:Keywords: duplication; glycoprotein; transmembrane protein  
F:1-41/Domain: signal sequence #status predicted <SIG>  
F:42-2039/Product: complement C3b/C4b receptor, membrane-bound #status predicted <MAT>  
F:43-99/Domain: complement factor H repeat homology <FH01>  
F:104-161/Domain: complement factor H repeat homology <FH02>  
F:166-232/Domain: complement factor H repeat homology <FH03>  
F:238-293/Domain: complement factor H repeat homology <FH04>  
F:297-353/Domain: complement factor H repeat homology <FH05>  
F:358-416/Domain: complement factor H repeat homology <FH06>  
F:421-487/Domain: complement factor H repeat homology <FH07>  
F:493-549/Domain: complement factor H repeat homology <FH08>  
F:554-611/Domain: complement factor H repeat homology <FH09>  
F:616-682/Domain: complement factor H repeat homology <FH09>





F:36-94/Domain: complement factor H repeat homology <FH01>  
F:98-158/Domain: complement factor H repeat homology <FH02>  
F:163-220/Domain: complement factor H repeat homology <FH03>  
F:225-283/Domain: complement factor H repeat homology <FH04>  
F:354-381/Domain: carboxyl-terminal propeptide #status predicted <PRO>  
F:5/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:353/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Ser) (in mature form)

Query Match 49.5%; Score 45; DB 1; Length 381;  
Best Local Similarity 61.5%; Pred. No. 3.6;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 KVELVGEPSIYC 17  
| : : : : |||  
Db 255 KGFTMIGESHIYC 267

## RESULT 11

decay-accelerating factor splice form 1 precursor - human  
N:Alternate names: decay-accelerating factor 5; decay-accelerating factor secreted form  
C:Species: Homo sapiens (man)  
C:Date: 05-Oct-1988 #sequence, revision 05-Oct-1988 #text\_change 29-Aug-1997  
C:Accession: A26359; A39702; S16187; S23138; A27258  
R:Caras, I.W.; Davitz, M.A.; Rhee, L.; Weddell, G.; Martin Jr., D.W.; Nussenzweig, V.  
Nature 325, 545-549, 1987  
A:Title: Cloning of decay-accelerating factor suggests novel use of splicing to generate  
A:Reference number: A26359; MUID:87115845  
A:Accession: A26359  
A:Molecule type: mRNA  
A:Residues: 1-440 <CAR>  
A:Cross-references: GB:M30142  
R:Emulonu, U.K.; Ravi, L.; Medof, M.E.  
Proc. Natl. Acad. Sci. U.S.A. 88, 4675-4679, 1991  
A:Title: Characterization of the decay-accelerating factor gene promoter region.  
A:Reference number: A39702; MUID:91271256  
A:Accession: A39702  
A:Molecule type: DNA  
A:Residues: 1-79, 'T', 81-104 <EMU>  
A:Cross-references: GB:M64356  
A:Note: The authors translated the codon AGT for residue 85 as Met  
R:Nakano, Y.; Sugita, Y.; Ishikawa, Y.; Choi, N.H.; Tobe, T.; Tomita, M.  
Biochim. Biophys. Acta 1074, 326-330, 1991  
A:Title: Isolation of two forms of decay-accelerating factor (DAF) from human urine.  
A:Reference number: S16187; MUID:91291869  
A:Accession: S16187  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 35-47 <BIO>  
R:Nakano, Y.; Sumida, K.; Kikuta, N.; Miura, N.H.; Tobe, T.; Tomita, M.  
Biochim. Biophys. Acta 1116, 235-240, 1992  
A:Title: Complete determination of disulfide bonds localized within the short consensus  
A:Reference number: S23138  
A:Accession: S23138  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 35-41; 65-68; 79-81; 93-103; 128-134; 143-145; 155-159; 162-168; 188-192; 203-204; 211  
R:Sugita, Y.; Negoro, T.; Matsuda, T.; Sakamoto, T.; Tomita, M.  
J. Biochem. 100, 143-150, 1986  
A:Title: Improved method for the isolation and preliminary characterization of human DAF  
A:Reference number: A27258; MUID:87008461  
A:Accession: A27258  
A:Molecule type: protein  
A:Residues: 35, 'X', 37, 'G', 39-51, 'P', 53-55, 'X', 57-58, 'X', 60-63 <SUG>  
A:Note: Gly-37 and Leu-38 were also found  
C:Genetics:  
A:Gene: GDB:DAF  
A:Cross-references: GDB:119088; OMIM:125240  
A:Map position: 1q32-1q32  
C:Superfamily: decay-accelerating factor; complement factor H repeat homology  
C:Keywords: alternative splicing; glycoprotein  
F:1-34/Domain: signal sequence #status predicted <SIG>

F:35-440/Product: decay-accelerating factor 1 #status predicted <MAT>  
F:36-94/Domain: complement factor H repeat homology <FH01>  
F:98-158/Domain: complement factor H repeat homology <FH02>  
F:163-220/Domain: complement factor H repeat homology <FH03>  
F:225-283/Domain: complement factor H repeat homology <FH04>  
F:35/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 49.5%; Score 45; DB 2; Length 440;  
Best Local Similarity 61.5%; Pred. No. 4.2;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 KVELVGEPSIYC 17  
| : : : : |||  
Db 255 KGFTMIGESHIYC 267

## RESULT 12

membrane cofactor protein - human  
C:Species: Homo sapiens (man)  
C:Date: 02-Jul-1996 #sequence, revision 02-Jul-1996 #text\_change 07-Aug-1998  
C:Accession: I57998  
R:Gervoni, F.; Fentichel, P.; Akhondi, C.; Hsi, B.L.; Rossi, B.  
Mol. Reprod. Dev. 34, 107-113, 1993  
A:Title: Characterization of a cDNA clone coding for human testis membrane cofactor p  
A:Reference number: I57998; MUID:93119658  
A:Accession: I57998  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-369 <RES>  
A:Cross-references: GB:S51940; NID:G262937; PID:G262938  
C:Superfamily: complement factor H repeat homology  
F:35-94/Domain: complement factor H repeat homology <FH01>  
F:162-223/Domain: complement factor H repeat homology <FH03>  
F:228-283/Domain: complement factor H repeat homology <FH02>

Query Match 49.5%; Score 45; DB 2; Length 369;  
Best Local Similarity 53.3%; Pred. No. 3.5;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 GRKVELVGEPSIYC 17  
| : : : : |||  
Db 196 GRPFFSLIGESHIYC 210

## RESULT 13

membrane cofactor protein precursor - human  
C:Species: Homo sapiens (man)  
C:Date: 30-Sep-1989 #sequence, revision 30-Sep-1989 #text\_change 10-Sep-1997  
C:Accession: S01896; A60765; I56188  
R:Rubin, D.M.; Liszewski, M.K.; Post, T.W.; Arce, M.A.; Le Beau, M.M.; Rebentisch, M.  
J. Exp. Med. 168, 181-194, 1988  
A:Title: Molecular cloning and chromosomal localization of human membrane cofactor pr  
A:Reference number: S01896; MUID:88286080  
A:Accession: S01896  
A:Molecule type: mRNA  
A:Residues: 1-384 <LUB>  
A:Cross-references: EMBL:Y00651; NID:G34504; PID:G34505  
A:Note: part of this sequence, including the amino end of the mature protein, was con  
R:Purcell, D.F.J.; Deacon, N.J.; Andrew, S.M.; McKenzie, I.F.C.  
Immunogenetics 31, 21-28, 1990  
A:Title: Human non-lineage antigen, CD46 (Huy-m5): purification and partial sequenci  
A:Reference number: A60765  
A:Accession: A60765  
A:Molecule type: protein  
A:Residues: 'X', 36-39, 'X', 41-58 <PUR>  
R:Cui, W.; Hourcade, D.; Post, T.; Greenland, A.C.; Atkinson, J.P.; Kumar, V.  
J. Immunol. 151, 4137-4146, 1993  
A:Title: Characterization of the promoter region of the membrane cofactor protein (CD

A:Reference number: 156188  
 A:Accession: 156188  
 C:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-34 <RES>  
 A:Cross-references: GB:S65879; NID:g425643  
 C:Genetics:  
 A:Gene: GDB:MCP  
 A:Cross-references: GDB:120169; OMIM:120920  
 A:Map position: 1q32-1q32  
 C:Function:  
 A:Description: membrane cofactor protein for the factor I-mediated cleavage of the comp  
 C:Superfamily: complement factor H repeat homology  
 C:Keywords: alternative splicing; glycoprotein; transmembrane protein  
 F:1-34/Domain: signal sequence #status predicted <SIG>  
 F:35-384/Product: membrane cofactor #status experimental <MAT>  
 F:35-94/Domain: complement factor H repeat homology <FH01>  
 F:99-157/Domain: complement factor H repeat homology <FH02>  
 F:162-223/Domain: complement factor H repeat homology <FH03>  
 F:228-283/Domain: complement factor H repeat homology <FH04>  
 F:329-351/Domain: transmembrane #status predicted <TM>

Query Match 49.5%; Score 45; DB 2; Length 384;  
 Best Local Similarity 53.3%; Pred. No. 3.6;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 3 GRKVFELVGEPSIYC 17  
 DB 196 GPDPSLIGESTIYC 210

#### RESULT 14

154479  
 membrane cofactor protein precursor, splice form pm5.1 - human  
 N:Alternate names: lymphocyte surface glycoprotein CD46  
 C:Species: Homo sapiens (man)  
 C:Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 07-Aug-1998  
 C:Accession: 154479  
 R:Purcell, D.F.; Russell, S.M.; Deacon, N.J.; Brown, M.A.; Hooker, D.J.; McKenzie, I.F.  
 A:Title: Alternatively spliced RNAs encode several isoforms of CD46 (MCP), a regulator o  
 A:Reference number: 154479; MUID:91267562  
 A:Accession: 154479  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-377 <RES>  
 A:Cross-references: GB:M58050; NID:g180136; PID:g180137  
 C:Genetics:  
 A:Gene: GDB:MCP  
 A:Cross-references: GDB:120169; OMIM:120920  
 A:Map position: 1q32-1q32  
 C:Function:  
 A:Description: for the factor I-mediated cleavage of the complement convertases  
 C:Superfamily: complement factor H repeat homology  
 C:Keywords: alternative splicing; glycoprotein; transmembrane protein  
 F:35-94/Domain: complement factor H repeat homology <FH01>  
 F:162-223/Domain: complement factor H repeat homology <FH03>  
 F:228-283/Domain: complement factor H repeat homology <FH02>  
 F:329-351/Domain: transmembrane #status predicted <TM>

Query Match 49.5%; Score 45; DB 2; Length 377;  
 Best Local Similarity 53.3%; Pred. No. 3.6;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 3 GRKVFELVGEPSIYC 17  
 DB 196 GPDPSLIGESTIYC 210

RESULT 15  
 G02913

sperm CD46 - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 07-Aug-1998  
 C:Accession: G02913  
 R:Hara, T.  
 Submitted to GenBank, March 1996  
 A:Reference number: H01942  
 A:Accession: G02913  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-349 <HR>  
 A:Cross-references: GB:D84105; NID:g1256700; PID:g1256701  
 C:Superfamily: complement factor H repeat homology  
 F:35-94/Domain: complement factor H repeat homology <FH01>  
 F:162-223/Domain: complement factor H repeat homology <FH03>  
 F:228-283/Domain: complement factor H repeat homology <FH02>

Query Match 49.5%; Score 45; DB 2; Length 349;  
 Best Local Similarity 53.3%; Pred. No. 3.3;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 3 GRKVFELVGEPSIYC 17  
 DB 196 GPDPSLIGESTIYC 210

Search completed: July 18, 1999, 06:07:04  
 Job time: 250 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 11:37:51 ; Search time 31.53 Seconds  
(without alignments)  
14.471 Million cell updates/sec

Title: US-09-142-043-5

Perfect score: 91

Sequence: 1 SGRKRVFELVGPSPSYC 17

Scoring table: BLOSUM62

Searched: 74019 seqs, 26840295 residues

Database : SwissProt\_36:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	91	100.0	2039	1 CRL_HUMAN	P17927 homo sapien
2	45	49.5	381	1 DAF_HUMAN	P08174 homo sapien
3	45	49.5	377	1 MCP_HUMAN	P15529 homo sapien
4	44	48.4	589	1 CAH_DUNSA	P54312 dunaliella
5	43.5	47.8	156	1 BODP_HUMAN	P36639 homo sapien
6	43	47.3	274	1 T2B1_HERAU	P25257 herpetosiph
7	43	47.3	274	1 T2E1_HERAU	P25260 herpetosiph
8	42	46.2	390	1 DAF1_MOUSE	P61475 mus musculu
9	42	46.2	340	1 DAF_PONPY	P94557 pongo pygma
10	42	46.2	158	1 DYRC_STRAU	P10167 staphylococ
11	42	46.2	165	1 DYRC_ECOLI	P05908 escherichia
12	41	45.1	168	1 DYR_BACSU	P11045 bacillus su
13	41	45.1	160	1 YB6A_SCHPO	P43791 haemophilus
14	41	45.1	1522	1 YB6A_SCHPO	P09750 schizosacch
15	40	44.0	1025	1 CR2_MOUSE	P19070 mus musculu
16	40	44.0	313	1 PCR_MOUSE	P15804 avena sativ
17	40	44.0	277	1 THTR_MYCLE	P05036 mycobacteri
18	40	44.0	277	1 THTR_MYCTU	P06538 mycobacteri
19	39	42.9	1061	1 DPOL_ABEI2	P05793 human adeno
20	38.5	42.3	291	1 CHIL1_PINTH	P41645 pinus contu
21	38.5	42.3	291	1 CHIL1_PINTH	P41645 pinus contu
22	38.5	42.3	291	1 CUT9_SCHPO	P41689 schizosacch
23	38.5	42.3	321	1 SOL1_YEAST	P50278 saccharomyc
24	38	41.8	582	1 CEAY_ECOLI	P09883 escherichia
25	38	41.8	162	1 DYR3_SALTY	P12833 salmonella
26	38	41.8	316	1 GSHB_ECOLI	P04423 escherichia
27	38	41.8	476	1 HGDA_ACIFE	P11569 acidammonoc
28	38	41.8	232	1 SMY_MOUSE	P13675 mus musculu
29	38	41.8	1548	1 UGG6_DROME	P03632 drosophila
30	37.5	41.2	384	1 YBRC_BACSU	P40407 bacillus su
31	37.5	41.2	384	1 STRH_STRGR	P09399 streptomyce
32	37.5	41.2	590	1 VG28_HSV1L	P00131 ictaluriid h
33	37	40.7	454	1 AATM_LUPAN	P26563 lupinus ang
34	37	40.7	1033	1 CR2_HUMAN	P20023 homo sapien
35	37	40.7	161	1 DYR_STAEP	P05908 staphylococ
36	37	40.7	180	1 MP17_LYMTS	P05908 staphylococ
37	37	40.7	401	1 PKR_SYNY3	P74421 synchocyst
38	37	40.7	230	1 VP26_NPVOP	P11037 synchocyst
39	37	40.7	1030	1 VPP1_CAEEL	P30628 caenorhabdi
40	37	40.7	453	1 YOTA_CAEEL	P34656 caenorhabdi
41	37	40.7	272	1 TYXD_BACSU	P39063 bacillus su
42	36.5	40.1	475	1 EX1_ECOLI	P04995 escherichia
43	36	39.6	467	1 ARLY_RANCA	P51464 rana catesb

## ALIGNMENTS

RESULT	1	44	36	39.6	353	1	CA26_HUMAN
CRL_HUMAN		45	36	39.6	647	1	COAT_AADV
ID	CRL_HUMAN						P12110 homo sapien
AC	P17927;						P24029 aleutian mi
DT	01-NOV-1990 (REL. 16, CREATED)						
DT	01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)						
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)						
DE	COMPLEMENT RECEPTOR TYPE 1 PRECURSOR (C3B/C4B RECEPTOR) (CD35						
DE	ANTIGEN)						
GN	CRI OR C3BR.						
OS	HOMO SAPIENS (HUMAN).						
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;						
OC	EUTHERIA; PRIMATES.						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RX	MEDLINE: 89035992.						
RA	KLICKSTEIN L.B., BARTOW T.J., MILETIC V., RABSON L.D., SMITH J.A.,						
RA	FEARON D.T.;						
RL	J. EXP. MED. 168:1699-1717(1988).						
RN	[2]						
RP	SEQUENCE OF 503-2039 FROM N.A.						
RX	MEDLINE: 87168191.						
RA	KLICKSTEIN L.B., WONG W.W., SMITH J.A., WEIS J.H., WILSON J.G.,						
RA	FEARON D.T.;						
RL	J. EXP. MED. 165:1095-1112(1987).						
RN	[3]						
RP	SEQUENCE OF 761-783; 831-845 AND 1179-1195 FROM N.A.						
RX	MEDLINE: 86067975.						
RA	WONG W.W., KLICKSTEIN L.B., SMITH J.A., WEIS J.H., FEARON D.T.;						
RA	PROC. NATL. ACAD. SCI. U.S.A. 82:7711-7715(1985).						
CC	-1- FUNCTION: CRI, PRESENT ON ERYTHROCYTES, LEUKOCYTES, GLOMERULAR						
CC	PODOCYTES, AND SPLENIC FOLLICULAR DENDRITIC CELLS. MEDIATES THE						
CC	BINDING BY THESE CELLS OF PARTICLES AND IMMUNE COMPLEXES THAT HAVE						
CC	ACTIVATED COMPLEMENT.						
CC	-1- SUBUNIT: MONOMER.						
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.						
CC	-1- POLYMORPHISM: CRI IS RESPONSIBLE FOR THE KNOX BLOOD GROUP SYSTEM.						
CC	-1- SEVEN SHORT CONSENSUS REPEATS (SCR) CONSTITUTE A LONG HOMOLOGOUS						
CC	REPEAT (LHR). THE N-TERMINAL TWO SCRS OF LHR-A CONTAINED A SITE						
CC	AND -C EACH HAD A SITE DETERMINING C3 SPECIFICITY.						
CC	-1- THIS IS THE SEQUENCE OF THE F ALLOTYPIC OF CRI.						
CC	-1- SIMILARITY: CONTAINS 30 SUSHI (SCR) REPEATS.						
DR	EMBL: M11569; G180995; -						
DR	EMBL: M11617; G180996; -						
DR	EMBL: Y00816; G30186; -						
DR	EMBL: X05309; G809019; -						
DR	PIR: A28507; A28507.						
DR	PIR: A24748; A24748.						
DR	PIR: B24748; B24748.						
DR	PIR: C24748; C24748.						
DR	HSSP: P08603; S03843.						
DR	MIM: 106620; -						
KW	COMPLEMENT PATHWAY; GLYCOPROTEIN; TRANSMEMBRANE; REPEAT; SIGNAL;						
KW	RECEPTOR; SUSHI; BLOOD GROUP ANTIGEN.						
FT	SIGNAL						
FT	CHAIN						
FT	DOMAIN						
FT	TRANSMEM						
FT	DOMAIN						
FT	MOD.RES						
FT	DOMAIN						
FT	REPEAT A.						
FT	REPEAT						

```
FT REPEAT 103 162 SUSHI A2.
FT REPEAT 165 233 SUSHI A3.
FT REPEAT 237 294 SUSHI A4.
FT REPEAT 296 354 SUSHI A5.
FT REPEAT 357 417 SUSHI A6.
FT REPEAT 420 488 SUSHI A7.
FT DOMAIN 492 938 7 x SUSHI
FT REPEAT 492 938 REPEATS, LONG HOMOLOGOUS
FT REPEAT 492 550 SUSHI B1.
FT REPEAT 553 612 SUSHI B2.
FT REPEAT 615 683 SUSHI B3.
FT REPEAT 687 744 SUSHI B4.
FT REPEAT 746 804 SUSHI B5.
FT REPEAT 807 867 SUSHI B6.
FT REPEAT 870 938 SUSHI B7.
FT DOMAIN 942 1388 7 x SUSHI (SCR) REPEATS, LONG HOMOLOGOUS
FT REPEAT 942 1000 REPEAT C.
FT REPEAT 1003 1062 SUSHI C1.
FT REPEAT 1065 1133 SUSHI C2.
FT REPEAT 1137 1194 SUSHI C3.
FT REPEAT 1196 1254 SUSHI C4.
FT REPEAT 1257 1317 SUSHI C5.
FT REPEAT 1320 1388 SUSHI C6.
FT DOMAIN 1395 1846 7 x SUSHI (SCR) REPEATS, LONG HOMOLOGOUS
FT REPEAT 1395 1453 SUSHI D1.
FT REPEAT 1456 1515 SUSHI D2.
FT REPEAT 1518 1586 SUSHI D3.
FT REPEAT 1590 1647 SUSHI D4.
FT REPEAT 1649 1707 SUSHI D5.
FT REPEAT 1710 1770 SUSHI D6.
FT REPEAT 1773 1841 SUSHI D7.
FT DOMAIN 1847 1966 2 x SUSHI (SCR) REPEATS.
FT REPEAT 1847 1905 SUSHI 1.
FT REPEAT 1908 1966 SUSHI 2.
FT DISULFID 43 86 BY SIMILARITY.
FT DISULFID 73 99 BY SIMILARITY.
FT DISULFID 104 145 BY SIMILARITY.
FT DISULFID 131 161 BY SIMILARITY.
FT DISULFID 166 215 BY SIMILARITY.
FT DISULFID 195 232 BY SIMILARITY.
FT DISULFID 238 280 BY SIMILARITY.
FT DISULFID 266 293 BY SIMILARITY.
FT DISULFID 297 340 BY SIMILARITY.
FT DISULFID 326 353 BY SIMILARITY.
FT DISULFID 358 400 BY SIMILARITY.
FT DISULFID 386 416 BY SIMILARITY.
FT DISULFID 421 470 BY SIMILARITY.
FT DISULFID 450 487 BY SIMILARITY.
FT DISULFID 493 536 BY SIMILARITY.
FT DISULFID 523 549 BY SIMILARITY.
FT DISULFID 554 595 BY SIMILARITY.
FT DISULFID 581 611 BY SIMILARITY.
FT DISULFID 616 665 BY SIMILARITY.
FT DISULFID 645 682 BY SIMILARITY.
FT DISULFID 688 730 BY SIMILARITY.
FT DISULFID 716 743 BY SIMILARITY.
FT DISULFID 747 790 BY SIMILARITY.
FT DISULFID 776 803 BY SIMILARITY.
FT DISULFID 808 850 BY SIMILARITY.
FT DISULFID 836 866 BY SIMILARITY.
FT DISULFID 871 920 BY SIMILARITY.
FT DISULFID 900 937 BY SIMILARITY.
FT DISULFID 943 986 BY SIMILARITY.
FT DISULFID 973 999 BY SIMILARITY.
FT DISULFID 1004 1045 BY SIMILARITY.
FT DISULFID 1031 1061 BY SIMILARITY.
FT DISULFID 1066 1115 BY SIMILARITY.
FT DISULFID 1095 1132 BY SIMILARITY.
FT DISULFID 1138 1180 BY SIMILARITY.
FT DISULFID 1166 1193 BY SIMILARITY.
FT DISULFID 1197 1240 BY SIMILARITY.
```

```
FT DISULFID 1226 1253 BY SIMILARITY.
FT DISULFID 1258 1300 BY SIMILARITY.
FT DISULFID 1286 1316 BY SIMILARITY.
FT DISULFID 1321 1370 BY SIMILARITY.
FT DISULFID 1350 1387 BY SIMILARITY.
FT DISULFID 1366 1439 BY SIMILARITY.
FT DISULFID 1426 1452 BY SIMILARITY.
FT DISULFID 1457 1498 BY SIMILARITY.
FT DISULFID 1484 1514 BY SIMILARITY.
FT DISULFID 1519 1568 BY SIMILARITY.
FT DISULFID 1548 1585 BY SIMILARITY.
FT DISULFID 1591 1633 BY SIMILARITY.
FT DISULFID 1619 1646 BY SIMILARITY.
FT DISULFID 1650 1693 BY SIMILARITY.
FT DISULFID 1679 1706 BY SIMILARITY.
FT DISULFID 1711 1753 BY SIMILARITY.
FT DISULFID 1739 1769 BY SIMILARITY.
FT DISULFID 1774 1823 BY SIMILARITY.
FT DISULFID 1803 1840 BY SIMILARITY.
FT DISULFID 1848 1891 BY SIMILARITY.
FT DISULFID 1877 1904 BY SIMILARITY.
FT DISULFID 1909 1952 BY SIMILARITY.
FT DISULFID 1938 1965 BY SIMILARITY.
FT CARBOHYD 56 56 POTENTIAL.
FT CARBOHYD 252 252 POTENTIAL.
FT CARBOHYD 410 410 POTENTIAL.
FT CARBOHYD 447 447 POTENTIAL.
FT CARBOHYD 509 509 POTENTIAL.
FT CARBOHYD 578 578 POTENTIAL.
FT CARBOHYD 702 702 POTENTIAL.
FT CARBOHYD 860 860 POTENTIAL.
FT CARBOHYD 897 897 POTENTIAL.
FT CARBOHYD 959 959 POTENTIAL.
FT CARBOHYD 1028 1028 POTENTIAL.
FT CARBOHYD 1152 1152 POTENTIAL.
FT CARBOHYD 1310 1310 POTENTIAL.
FT CARBOHYD 1481 1481 POTENTIAL.
FT CARBOHYD 1504 1504 POTENTIAL.
FT CARBOHYD 1534 1534 POTENTIAL.
FT CARBOHYD 1540 1540 POTENTIAL.
FT CARBOHYD 1605 1605 POTENTIAL.
FT CARBOHYD 1763 1763 POTENTIAL.
FT CARBOHYD 1908 1908 POTENTIAL.
SQ SEQUENCE 2039 AA; 223589 MW; 666F9033 CRC32;

Query Match 100.0%; Score 91; DB 1; Length 2039;
Best Local Similarity 100.0%; Pred. No. 1.9e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SGGKRVFELVGPSTYC 17
Db 199 SGGKRVFELVGPSTYC 215

RESULT 2
ID DAF_HUMAN STANDARD; PRT; 381 AA.
AC P08174; P09679;
DT 01-AUG-1988 (REL. 08, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE COMPLEMENT DECAV-ACCELERATING FACTOR PRECURSOR (CD55).
OS DAF OR CD55.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 87115845.
RA CARAS I.W., DAVITZ M.A., RHEE I., WEDDELL G., MARTIN D.W. JR.,
RA NOSSENZWEIG V.;
RL NATURE 325:545-549(1987).
```

CC	DAF-1	AAE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.
CC	-1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.	
CC	-1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION (RCA) FAMILY.	
CC	-1- DATABASE: NAME=PROW; NOTE=CD guide CD55 entry; WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd55.htm".	
CC	EMBL; M31516; G181468; -	
DR	EMBL; M30142; G181465; -	
DR	EMBL; M15799; G181463; -	
DR	EMBL; M64653; G181476; -	
DR	EMBL; M64356; G181476; JOINED.	
DR	EMBL; S72858; G639600; -	
DR	PIR; B26359; B26359.	
DR	PIR; A26359; A26359.	
DR	PIR; S16187; S16187.	
DR	PIR; A39101; A39101.	
DR	PIR; S23138; S23138.	
DR	HSSP; P08603; LHFI.	
DR	MM; 125240; -	
KV	COMPLETE PATHWAY: PLASMA; GLYCOPROTEIN; MEMBRANE; REPEAT;	
KV	ALTERNATIVE SPLICING; GPI-ANCHOR; SIGNAL; SUSH1; POLYMORPHISM;	
KW	BLOOD GROUP ANTIGEN.	
FT	SIGNAL	1 34
FT	CHAIN	35 353
FT	PROPEP	354 381
FT	DOMAIN	35 284
FT	REPEAT	35 95
FT	REPEAT	97 159
FT	REPEAT	162 221
FT	REPEAT	224 284
FT	REPEAT	287 356
FT	DOMAIN	36 81
FT	DISULFID	65 94
FT	DISULFID	98 145
FT	DISULFID	129 158
FT	DISULFID	163 204
FT	DISULFID	190 220
FT	DISULFID	225 267
FT	DISULFID	253 283
FT	CARBOHYD	95 95
FT	LIPID	353 353
FT	VARIANT	52 52
FT	VARIANT	52 52
FT	VARIANT	82 82
FT	VARIANT	199 199
FT	VARIANT	227 227
FT	VARIANT	362 381
FT	VARSPLIC	
FT	POTENTIAL.	
FT	GPI-ANCHOR.	
FT	R -> L (IN TC(B) ANTIGEN).	
FT	R -> P (IN TC(C) ANTIGEN).	
FT	L -> R (IN WEG(A) ANTIGEN).	
FT	S -> L (IN DR(A-) ANTIGEN).	
FT	A -> P (IN CR(A-) ANTIGEN).	
FT	HTCFTLGLTGLTIVMGLT -> SRPYTQAGMWCDSRSL	
FT	OSRTGFERSRFSHSLSSSMYRAHFVHDFRAMDASNNGLA	
FT	DLAKELRRKTYOYRLELVS (IN DAF-1).	
FT	T -> I (IN REF. 1).	
FT	S -> M (IN REF. 2).	
FT	CONFLICT	80 80
FT	CONFLICT	85 85
FT	CONFLICT	381 AA; 41388 MW; 114271AB CRC32;
SO	SEQUENCE	
QY	Query Match	49.5%; Score 45; DB 1; Length 381;
QY	Best Local Similarity	61.5%; Pred. No. 2.5;
QY	Matches	8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY	5 KVFELVGEPSIYC 17	
QY	1 1 ::::	
QY	255 KGTMTGEHSIYC 267	
RESULT	3	
ID	MCP_HUMAN	STANDARD; PRT; 377 AA.
AC	P15529;	
DT	01-APR-1990 (REL. 14, CREATED)	
DT	01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)	
DT	15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)	
DE	MEMBRANE COFACTOR PROTEIN PRECURSOR (CD46 ANTIGEN) (TROPHOBLAST	
DE	LEUCOCYTE COMMON ANTIGEN) (TLX).	
DE	MCP.	

OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 35-58.  
 RX MEDLINE; 88286080.  
 RA LOBLIN D.M., LISZESKI M.K., POST T.W., ARCE M.A., LE BEAU M.M.,  
 RA REBENTISCH M.B., LEMONS R.S., SEVA T., ATKINSON J.P.;  
 RL J. EXP. MED. 168:181-194(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-TESTIS;  
 RX MEDLINE; 93119658.  
 RA CERVOINI F., FENICHEL P., AKHOUNDI C., HSI B.L., ROSSI B.;  
 RL MOL. REPROD. DEV. 34:107-113(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RX MEDLINE; 91267562.  
 RA PURCELL D.F., RUSSELL S.M., DEACON N.J., BROWN M.A., HOOKER D.J.,  
 RA MCKENZIE I.F.;  
 RL IMMUNOGENETICS 33:335-344(1991).  
 RN [4]  
 RP SEQUENCE OF 1-34 FROM N.A.  
 RX MEDLINE; 94014356.  
 RA COI W., HOURCADE D., POST T., GREENLUND A.C., ATKINSON J.P.,  
 RA KUMAR V.;  
 RL J. IMMUNOL. 151:4137-4146(1993).  
 RN [5]  
 RP ALTERNATIVE SPLICING.  
 RX MEDLINE; 92289809.  
 RA RUSSELL S.M., SPARROW R.L., MCKENZIE I.F.C., PURCELL D.F.J.;  
 RL EUR. J. IMMUNOL. 22:1513-1518(1992).  
 CC -1- FUNCTION: MCP IS AN IMPORTANT MEMBRANE PROTEIN FOR PROTECTING HOST  
 CC CELLS FROM DAMAGE BY COMPLEMENT. MCP POSSESSES COFACTOR ACTIVITY  
 CC FOR INACTIVATION OF C3B AND C4B BY SERUM FACTOR I. THIS WOULD  
 CC ALLOW MCP TO REGULATE THE COMPLEMENT SYSTEM AT THE LEVEL OF THE C3  
 CC CONVERTASES. MAY BE INVOLVED IN THE FUSION OF THE SPERMATOZOA WITH  
 CC THE OOCYTE DURING FERTILIZATION. MAY ACT BY PREVENTING COMPLEMENT  
 CC ACTIVATION AT THE FETO-MATERNAL INTERFACE ON THE  
 CC SYNCTIOTROPHOBLAST LAYER OF PLACENTA.  
 CC SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED BY ALMOST ALL CELLS EXCEPT  
 CC ERYTHROCYTES AND SOME BONE MARROW CELLS.  
 CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED IN THE SER/THR-RICH DOMAIN  
 CC (PROBABLE).  
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 14 FORMS ARE PRODUCED BY  
 CC ALTERNATIVE SPLICING. THE FORM SHOWN IS THE TRANSCRIPT A. THE  
 CC GAMMA ISOFORM (TRANSCRIPTS A AND B) IS PREFERENTIALLY EXPRESSED IN  
 CC EBV-B CELLS AND LEUKEMIC CELLS. THE 66 KD ALPHA ISOFORM AND THE 56  
 CC KD BETA ISOFORM FOUND IN ALL TISSUES EXCEPT SPERM CORRESPOND TO  
 CC THE TRANSCRIPTS C AND D, AND THE TRANSCRIPTS E AND F RESPECTIVELY.  
 CC THE EXON 9 IS SPECIFICALLY DELETED IN SOME PLACENTAE. THE EPITLON  
 CC ISOFORM CORRESPONDS TO THE TRANSCRIPTS I AND J. SPERMATOZOA DELTA  
 CC ISOFORM (TRANSCRIPT N) DELETE BOTH EXONS 12 AND 13. ALL TISSUES  
 CC DIFFERENTIALLY SPLICED EXON 13.  
 CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.  
 CC -1- DATABASE: NAME=PROV; NOTE=CD guide CD46 entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd46.htm".  
 DR EMBL: Y00651; G34505; -;  
 DR EMBL: S51940; G262938; -;  
 DR EMBL: M58050; G180137; -;  
 DR EMBL: A18585; G512457; -;  
 DR EMBL: S65879; E91388; -;  
 DR PIR: S01896; S01896.  
 DR MTN: 120920; -;  
 KW COMPLEMENT PATHWAY; GLYCOPROTEIN; TRANSMEMBRANE; REPEAT; SIGNAL;  
 KW SUSHI; ALTERNATIVE SPLICING.  
 FT SIGNAL 1 34  
 FT CHAIN 35 377 MEMBRANE COFACTOR PROTEIN.  
 FT DOMAIN 35 328 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 329 351 POTENTIAL.  
 FT CYTOPLASMIC 377 (POTENTIAL).  
 FT DOMAIN 35 284 4 X SUSHI (SCR) REPEATS.

FT REPEAT 35 95 SUSHI 1.  
 FT REPEAT 98 158 SUSHI 2.  
 FT REPEAT 161 224 SUSHI 3.  
 FT REPEAT 227 284 SUSHI 4.  
 FT DOMAIN 287 311 SER/THR-RICH.  
 FT DISULFID 35 80 BY SIMILARITY.  
 FT DISULFID 64 94 BY SIMILARITY.  
 FT DISULFID 99 141 BY SIMILARITY.  
 FT DISULFID 127 157 BY SIMILARITY.  
 FT DISULFID 162 210 BY SIMILARITY.  
 FT DISULFID 191 223 BY SIMILARITY.  
 FT DISULFID 228 270 BY SIMILARITY.  
 FT DISULFID 256 283 BY SIMILARITY.  
 FT CARBOHYD 83 83 POTENTIAL.  
 FT CARBOHYD 114 114 POTENTIAL.  
 FT CARBOHYD 273 273 POTENTIAL.  
 FT VARSPPLIC 286 300 MISSING (IN A 2ND FORM).  
 FT VARSPPLIC 340 352 MISSING (IN TRANSCRIPT M).  
 FT VARSPPLIC 353 361 YLQRRKKG -> DIFKGGRRGKQWELNMPLTRNLPLO  
 FT VARSPPLIC 362 377 OSREA (IN TRANSCRIPT M).  
 FT VARSPPLIC 362 377 TYLDETHREVEFTSL -> KADGGAEVATYQTKSTPAEQ  
 FT VARSPPLIC 362 377 RG (IN TRANSCRIPTS B, D, F, H, J, AND L).  
 FT VARSPPLIC 362 377 MISSING (IN TRANSCRIPTS M AND N).  
 FT VARSPPLIC 362 377 I -> IGVQWELNMPLTRNLPLOOSREA (IN  
 FT VARSPPLIC 362 377 TRANSCRIPT N).  
 SO SEQUENCE 377 AA; 42247 MW; 7FE9E38E CRC32;

Query Match 49.5%; Score 45; DB 1; Length 377;  
 Best Local Similarity 53.3%; Pred. No. 2.4;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 3 GRKVELYGEPSIYC 17  
 DB 196 GPDPELSIGESTIYC 210

RESULT 4  
 CAH\_DUNSA STANDARD; PRT; 589 AA.  
 AC P54212;  
 DT 01-OCT-1996 (REL. 34, CREATED)  
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE CARBONIC ANHYDRASE (EC 4.2.1.1) (CARBONATE DEHYDRATASE).  
 GN DCA.  
 OS DUNALIELLA SALINA.  
 OC EUKARYOTA; PLANTA; PHYCOPHYTA; CHLOROPHYTA (GREEN ALGAE);  
 OC CHLOROPHYCEAE; VOLVOCALES; DUNALIELLACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 96279304.  
 RA FISHER M., GORKHMAN I., PICK U., ZAMIR A.;  
 RL J. BIOL. CHEM. 271:17718-17723(1996).  
 CC -1- FUNCTION: REVERSIBLE HYDRATATION OF CARBON DIOXIDE.  
 CC -1- CATALYTIC ACTIVITY: H(2)CO(3) - CO(2) + H(2)O.  
 CC -1- INDUCTION: BY SALT.  
 CC -1- SIMILARITY: BELONGS TO THE CARBONIC ANHYDRASE FAMILY.  
 DR EMBL: U53811; G1431878; -;  
 KW LYSER; ZINC.  
 FT DOMAIN 390 589 ZINC CATALYTIC (BY SIMILARITY).  
 FT METAL 418 418 ZINC CATALYTIC (BY SIMILARITY).  
 FT METAL 420 420 ZINC CATALYTIC (BY SIMILARITY).  
 FT METAL 440 440 ZINC CATALYTIC (BY SIMILARITY).  
 FT METAL 440 440 ZINC CATALYTIC (BY SIMILARITY).  
 SO SEQUENCE 589 AA; 64257 MW; 2D375843 CRC32;

Query Match 48.4%; Score 44; DB 1; Length 589;  
 Best Local Similarity 66.7%; Pred. No. 5.7;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

DB 309 GRKVELVGEPSN 320

```

RESULT 5
BODP_HUMAN STANDARD: PRT: 156 AA.
ID P36639;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE 7-8-DIHYDRO-8-OXOGUANINE TRIPHOSPHATASE (EC 3.1.6.-) (8-OXO-DGTPASE).
GN MTH1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE: 94043152.
RA SAKUMI K., FURUCHI M., TSUZUKI T., KAKUMA T., KAWABATA S.-I.,
RA MAKI H., SEKIGUCHI M.;
RL J. BIOL. CHEM. 268:23524-23530(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-PLACENTA PERIPHERAL BLOOD;
RX MEDLINE: 95229148.
RA FURUCHI M., YOSHIDA M.C., ODA H., TAJIRI T., NAKABEPPU Y.,
RA TSUZUKI T., SEKIGUCHI M.;
RL GENOMICS 24:485-490(1994).
CC -!- FUNCTION: ANTIMUTAGENIC. RESPONSIBLE FOR PREVENTING
MISINCORPORATION OF 8-OXO-DGTP INTO DNA THUS PREVENTING A:T TO C:G
TRANSVERSIONS.
CC -!- CATALYTIC ACTIVITY: 8-OXO-DGTP + H(2)O = 8-OXO-DGMP +
PYROPHOSPHATE.
CC -!- PTM: THE N-TERMINAL IS BLOCKED.
CC -!- SIMILARITY: TO PROTEINS WITH A CORE MUTT DOMAIN.
DR EMBL: D1681; G452589; -
DR EMBL: D38594; G1405350; -
DR EMBL: D38592; G1405350; JOINED.
DR EMBL: D38593; G1405350; JOINED.
DR PIR: A4886; A48866.
DR MIM: 600312; -
DR PROSITE: PS00893; MUTT; 1.
KW HYDROLASE.
FT DOMAIN 33 MUTT-LIKE.
SQ SEQUENCE 156 AA; 17951 MW; ABFF97F0 CRC32;

```

Query Match 47.8%; Score 43.5; DB 1; Length 156;  
 Best Local Similarity 45.0%; Pred. No. 1.8;  
 Matches 9; Conservative 3; Mismatches 3; Indels 5; Gaps 1;

OY 3 GRKVELVGEPS-----IYC 17  
 I:|||||  
 DB 68 GQVFEFVEGPELMDVHVC 87

```

RESULT 6
T2BL_HERAU STANDARD: PRT: 274 AA.
ID P25257;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE TYPE II RESTRICTION ENZYME HG1BI (EC 3.1.21.4) (ENDONUCLEASE HG1BI)
DE (R.HG1BI).
GN HG1BI.
OS HERPETOSEIPHON AURANTIACUS (HERPETOSEIPHON GIGANTEUS).
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA;
OC NONPHOTOSYNTHETIC, NONFRUITING GLIDING; BEGGIATOACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HPG5;

```

```

RX MEDLINE: 91286195.
RA DUESTERHOEFT A., ERDMANN D., KROEGER M.;
RL NUCLEIC ACIDS RES. 19:3207-3211(1991).
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE: 95331596.
RA KROEGER M., BLUM E., DEPPE E., DUESTERHOEFT A., ERDMANN D., KILZ S.,
RA MEYER-ROGE S., MOESTL D.;
RL GENE 157:43-47(1995).
CC -!- CATALYTIC ACTIVITY: RECOGNIZES THE DOUBLE-STRANDED SEQUENCE GCMCC
AND CLEAVES AFTER G-1.
DR EMBL: X55137; G43479; -
DR PIR: S22308; S22308.
DR REBASE: RB00548; HG1BI.
KW HYDROLASE; ENDONUCLEASE; NUCLEASE; RESTRICTION SYSTEM.
SQ SEQUENCE 274 AA; 31188 MW; F523960E CRC32;

```

Query Match 47.3%; Score 43; DB 1; Length 274;  
 Best Local Similarity 50.0%; Pred. No. 3.8;  
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 3 GRKVELVGEPSIT 16  
 I:~:~:|||||  
 DB 237 GQFWSLIGEPSTY 250

```

RESULT 7
T2EL_HERAU STANDARD: PRT: 274 AA.
ID T2EL_HERAU
AC P25260;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE TYPE II RESTRICTION ENZYME HG1EI (EC 3.1.21.4) (ENDONUCLEASE HG1EI)
DE (R.HG1EI).
GN HG1EI.
OS HERPETOSEIPHON AURANTIACUS (HERPETOSEIPHON GIGANTEUS).
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA;
OC NONPHOTOSYNTHETIC, NONFRUITING GLIDING; BEGGIATOACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HPG24;
RA ERDMANN D., KROEGER M.;
RL SUBMITTED (NOV-1990) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE: 95331596.
RA KROEGER M., BLUM E., DEPPE E., DUESTERHOEFT A., ERDMANN D., KILZ S.,
RA MEYER-ROGE S., MOESTL D.;
RL GENE 157:43-47(1995).
CC -!- CATALYTIC ACTIVITY: RECOGNIZES THE DOUBLE-STRANDED SEQUENCE GCMCC
AND CLEAVES AFTER G-1.
DR EMBL: X55142; G43472; -
DR REBASE: RB00554; HG1EI.
KW HYDROLASE; ENDONUCLEASE; NUCLEASE; RESTRICTION SYSTEM.
SQ SEQUENCE 274 AA; 31325 MW; 7F4A74E5 CRC32;

```

Query Match 47.3%; Score 43; DB 1; Length 274;  
 Best Local Similarity 50.0%; Pred. No. 3.8;  
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 3 GRKVELVGEPSIT 16  
 I:~:~:|||||  
 DB 237 GQFWSLIGEPSTY 250

```

RESULT 8
DAFL_MOUSE STANDARD: PRT: 390 AA.
ID DAFL_MOUSE
AC Q61475; Q61397;
DT 01-NOV-1997 (REL. 35, CREATED)

```

DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE COMPLEMENT DECAY-ACCELERATING FACTOR, GPI-ANCHORED PRECURSOR  
DE (DAF-GPI).  
GN DAF1.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=TESTIS;  
RX MEDLINE: 95403982.  
RA SPICER A.P., SELDIN M.F., GENDLER S.J.;  
RL J. IMMUNOL. 155:3079-3091(1995).  
RN [2]  
RP SEQUENCE OF 7-390 FROM N.A.  
RC STRAIN=BALB/C; TISSUE=SPLEEN;  
RX MEDLINE: 96362213.  
RA FUKUDA Y., YASUI A., OKADA N., OKADA H.;  
RL INT. IMMUNOL. 8:379-385(1996).  
CC -1- FUNCTION: PROTECTION OF CELLS FROM COMPLEMENT-MEDIATED DAMAGE (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.  
CC -1- TISSUE SPECIFICITY: BRAIN, SECRETORY EPITHELIA, SKELETAL MUSCLE, LIVER, TESTES, THYMUS, SPLEEN AND LYMPH NODE.  
CC -1- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE ACTIVE SITE ON SCR3 (BY SIMILARITY).  
CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.  
CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION (RCA) FAMILY.  
DR EMBL: L41366; G886335; .  
DR EMBL: D63679; D1010476; .  
DR MGD: MGI:104850; DAF1.  
KM COMPLEMENT PATHWAY. GLYCOPROTEIN; MEMBRANE; REPEAT; GPI-ANCHOR;  
KM SIGNAL; SUSHI.  
FT SIGNAL 1 34  
FT CHAIN 35 361  
FT PROPER 362 390  
FT DOMAIN 35 285  
FT REPEAT 35 95  
FT REPEAT 97 159  
FT REPEAT 162 221  
FT REPEAT 224 285  
FT DOMAIN 288 364  
FT DISULFID 65 94  
FT DISULFID 98 145  
FT DISULFID 129 158  
FT DISULFID 163 204  
FT DISULFID 190 220  
FT DISULFID 225 267  
FT DISULFID 253 284  
FT CARBOHYD 187 187  
FT CARBOHYD 262 262  
FT LIPID 361 361  
FT CONFLICT 7 7  
FT CONFLICT 9 9  
FT CONFLICT 83 83  
FT CONFLICT 91 91  
FT CONFLICT 135 135  
FT CONFLICT 173 173  
FT CONFLICT 180 180  
SQ SEQUENCE 390 AA; 42618 MW; ADBDB34 CRC32;  
Query Match 46.2%; Score 42; DB 1; Length 390;  
Best Local Similarity 69.2%; Pred. No. 8.2;  
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

RESULT 9  
ID DAF\_PONPY STANDARD; PRT; 340 AA.  
AC P49457;  
DT 01-FEB-1996 (REL. 33, CREATED)  
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE COMPLEMENT DECAY-ACCELERATING FACTOR (CD55) (FRAGMENT).  
GN DAF-OR CD55.  
OS PONGO PYGMAEUS (ORANGUTAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 94110622.  
RA NICKELS M.W., ALVAREZ J.I., LOBLIN D.M., ATKINSON J.P.;  
RL J. IMMUNOL. 152:676-685(1994).  
CC -1- FUNCTION: THIS PROTEIN RECOGNIZES C4B AND C3B FRAGMENTS THAT CONDENSE WITH CELL-SURFACE HYDROXYL OR AMINO GROUPS WHEN NASCENT C4B AND C3B ARE LOCALLY GENERATED DURING C4 AND C3 ACTIVATION.  
CC INTERACTION OF DAF WITH CELL-ASSOCIATED C4B AND C3B POLYPEPTIDES INTERFERES WITH THEIR ABILITY TO CATALYZE THE CONVERSION OF C2 AND FACTOR B TO ENZYMATICALLY ACTIVE C2A AND BB AND THEREBY PREVENTS THE FORMATION OF C4B2A AND C3BBB, THE AMPLIFICATION CONVERTASES OF THE COMPLEMENT CASCADE (BY SIMILARITY).  
CC -1- SUBUNIT: MONOMER (MAJOR FORM) AND NON-DISULFIDE-LINKED, COVALENT HOMODIMER (MINOR FORM).  
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.  
CC -1- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE ACTIVE SITE ON SCR3 (BY SIMILARITY).  
CC -1- PTM: THE SER/THR-RICH DOMAIN IS HEAVILY O-GLYCOSYLATED.  
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF DAF (DAF-2, SHOWN HERE, AND DAF-1) ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.  
CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.  
CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION (RCA) FAMILY.  
DR EMBL: S67775; G459390; .  
KM COMPLEMENT PATHWAY. PLASMA; GLYCOPROTEIN; MEMBRANE; REPEAT;  
KM ALTERNATIVE SPLICING; GPI-ANCHOR; SUSHI.  
FT NON\_TER 1 312  
FT CHAIN 1 312  
FT PROPER 313 340  
FT DOMAIN <1 243  
FT REPEAT <1 54  
FT REPEAT 56 118  
FT REPEAT 121 180  
FT REPEAT 183 243  
FT DOMAIN 246 315  
FT DISULFID 24 53  
FT DISULFID 57 104  
FT DISULFID 88 117  
FT DISULFID 122 163  
FT DISULFID 149 179  
FT DISULFID 184 226  
FT DISULFID 212 242  
FT CARBOHYD 54 54  
FT CARBOHYD 107 107  
FT LIPID 312 312  
SQ SEQUENCE 340 AA; 37180 MW; 5CACC18F8 CRC32;  
Query Match 46.2%; Score 42; DB 1; Length 340;  
Best Local Similarity 53.8%; Pred. No. 7.1;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

RESULT 10  
ID DYR\_STAU STANDARD; PRT; 158 AA.  
AC P10167;  
DT 01-MAR-1989 (REL. 10, CREATED)  
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
DE DIHYDROFOLATE REDUCTASE TYPE I (EC 1.5.1.3).  
OS STAPHYLOCOCCUS AUREUS.  
OC PROKARYOTA; FIRMICUTES; COCCI; MICROCOCCACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 25923;  
RX MEDLINE; 93371025.  
RA DALE G.E., THEN R.L., STUEBER D.;  
RL ANTIMICROB. AGENTS CHEMOTHER. 37:1400-1405(1993).  
RN [2]  
RP SEQUENCE OF 1-35.  
RC STRAIN=157/4696;  
RX MEDLINE; 89078604.  
RA HARTMAN P.G., STAHLI M., KOCHER H.P., THEN R.L.;  
RL FEBS LETT. 242:157-160(1988).  
CC -1- CATALYTIC ACTIVITY: 5,6,7,8-TETRAHYDROFOLATE + NADP(+) =  
7,8-DIHYDROFOLATE + NADPH.  
CC -1- PATHWAY: ESSENTIAL STEP FOR DE NOVO GLYCINE AND PURINE SYNTHESIS,  
DNA PRECURSOR SYNTHESIS, AND FOR THE CONVERSION OF DUMP TO DIMP.  
CC -1- THERE ARE TWO DHFR ISOZYMES IN S.AUREUS, THIS ONE IS CHROMOSOMAL  
AND IS SENSITIVE TO TRIMETHOPRIM.  
CC -1- SIMILARITY: STRONG TO OTHER TYPE I DIHYDROFOLATE REDUCTASES.  
DR EMBL: Z16422; G49313; -.  
DR PIR: S02008; S02008.  
DR PIR: S32014; S32014.  
DR HSSP: P00379; 1DHI.  
DR PROSITE: PS00075; DHFR: 1.  
KW OXIDOREDUCTASE; NADP: ONE-CARBON METABOLISM.  
FT INIT MET 0  
SQ SEQUENCE 158 AA; 18120 MW; 1347E99F CRC32;

Query Match 46.2%; Score 42; DB 1; Length 158;  
Best Local Similarity 63.6%; Pred. No. 3.2;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 3 GRKFELVGP 13  
||| ||| :|:  
Db 43 GRKFESIGRP 53

RESULT 11  
ID DYRC\_ECOLI STANDARD; PRT; 165 AA.  
AC 059408;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DE DIHYDROFOLATE REDUCTASE TYPE XIII (EC 1.5.1.3).  
GN DHFRXIII.  
OS ESCHERICHIA COLI.  
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
ENTEROBACTERIACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RA3.2;  
SUBMITTED (AUG-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- CATALYTIC ACTIVITY: 5,6,7,8-TETRAHYDROFOLATE + NADP(+) =  
7,8-DIHYDROFOLATE + NADPH.  
CC -1- PATHWAY: ESSENTIAL STEP FOR DE NOVO GLYCINE AND PURINE SYNTHESIS,  
DNA PRECURSOR SYNTHESIS, AND FOR THE CONVERSION OF DUMP TO DIMP.  
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
CC -1- SIMILARITY: STRONG TO OTHER TYPE I DIHYDROFOLATE REDUCTASES.  
DR EMBL: Z50802; G950044; -.  
GN PROSITE: PS00075; DHFR: \*1.

KW OXIDOREDUCTASE; NADP; TRIMETHOPRIM RESISTANCE;  
KW METHOTREXATE RESISTANCE; ONE-CARBON METABOLISM.  
SQ SEQUENCE 165 AA; 17925 MW; C8BC9B76 CRC32;

Query Match 46.2%; Score 42; DB 1; Length 165;  
Best Local Similarity 63.6%; Pred. No. 3.4;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 3 GRKFELVGP 13  
||| ||| :|:  
Db 49 GRKFESIGRP 59

RESULT 12  
ID DYR\_BACSU STANDARD; PRT; 168 AA.  
AC P11045;  
DT 01-JUL-1989 (REL. 11, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE DIHYDROFOLATE REDUCTASE (EC 1.5.1.3).  
GN DFR.  
OS BACILLUS SUBTILIS.  
OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-8 AND 166-168.  
RC STRAIN=M112;  
RX MEDLINE; 88284366.  
RA IWAKURA M., KAWATA M., TSUDA K., TANAKA T.;  
RL GENE 64:9-20(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168 / MARBURG;  
RA SOROKIN A.V., AZEVEDO V., ZUMSTEIN E., GALLERON N., EHRLICH S.D.,  
RA SERROR P.;  
RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- CATALYTIC ACTIVITY: 5,6,7,8-TETRAHYDROFOLATE + NADP(+) =  
7,8-DIHYDROFOLATE + NADPH.  
CC -1- PATHWAY: ESSENTIAL STEP FOR DE NOVO GLYCINE AND PURINE SYNTHESIS,  
DNA PRECURSOR SYNTHESIS, AND FOR THE CONVERSION OF DUMP TO DIMP.  
CC -1- SIMILARITY: STRONG TO OTHER TYPE I DIHYDROFOLATE REDUCTASES.  
DR EMBL: M20012; G143742; -.  
DR EMBL: L77246; G1256641; -.  
DR EMBL: Z99115; E1183628; -.  
DR PIR: J70291; RDBSD.  
DR HSSP: P00379; 1DHI.  
DR SUBTILIST; BG10795; DFR.  
DR PROSITE: PS00075; DHFR: 1.  
KW OXIDOREDUCTASE; NADP: ONE-CARBON METABOLISM.  
FT VARIANT 165  
SQ SEQUENCE 168 AA; 19176 MW; 10142040 CRC32;

Query Match 45.1%; Score 41; DB 1; Length 168;  
Best Local Similarity 63.6%; Pred. No. 5.1;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 3 GRKFELVGP 13  
||| ||| :|:  
Db 43 GRKFESIGRP 53

RESULT 13  
ID DYR\_HAEIN STANDARD; PRT; 160 AA.  
AC P43791;  
DT 01-NOV-1995 (REL. 32, CREATED)  
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE DIHYDROFOLATE REDUCTASE (EC 1.5.1.3).  
GN FOLA OR FOLH OR H10899.  
OS HAEMOPHILUS INFLUENZAE.

OC PROKARYOTA: GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
OC PASTEURILLACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-RD / KW20:  
RX MEDLINE: 95350630.  
RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,  
RA KERNAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,  
RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,  
RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODER A., KELLEY J.M.,  
RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLUM E., COTTON M.D.,  
RA UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUNDER D.M., BRANDON R.C.,  
RA FINE L.D., FITCHMAN J.L., FUHRMANN J.L., GEORGAGEN N.S.M.,  
RA GENEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,  
RA VENTER J.C.;  
RL SCIENCE 269:496-512(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ISOLATES R906, R1042, AND R1047;  
RL DE GROOT R.;  
CC SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- CATALYTIC ACTIVITY: 5.6.7.8-TETRAHYDROLYATE + NADP(+) =  
CC 7.8-DIHYDROLYATE + NADPH.  
CC -1- PATHWAY: ESSENTIAL STEP FOR DE NOVO GLICINE AND PURINE SYNTHESIS,  
CC DNA PRECURSOR SYNTHESIS, AND FOR THE CONVERSION OF DIMP TO DIMP.  
CC -1- ISOLATES R906, R1042, AND R1047 ARE TRIMETHOPRIM-RESISTANT.  
CC -1- SIMILARITY: STRONG TO OTHER TYPE I DIHYDROLYATE REDUCTASES.  
DR EMBL: U32772: G1573919: -.  
DR EMBL: X84207: G665643: -.  
DR EMBL: X84205: G665645: -.  
DR EMBL: X84206: G665647: -.  
DR TIGR: H10899: -.  
DR PROSITE: PS00075; DHFR; 1.  
KW OXIDOREDUCTASE; NADP: ONE-CARBON METABOLISM: TRIMETHOPRIM RESISTANCE.  
FT VARIANT 13 13 N -> S (IN ISOLATES R1042 AND R1047).  
FT VARIANT 21 21 M -> I (IN ISOLATE R1047).  
FT VARIANT 34 34 P -> A (IN ISOLATE R1047).  
FT VARIANT 56 56 P -> A (IN ISOLATE R906).  
FT VARIANT 67 67 L -> P (IN ISOLATE R1042).  
FT VARIANT 69 69 E -> K (IN ISOLATE R1042).  
FT VARIANT 74 74 I -> V (IN ISOLATE R1042).  
FT VARIANT 77 77 D -> N (IN ISOLATE R1042).  
FT VARIANT 79 79 F -> L (IN ISOLATE R1042).  
FT VARIANT 95 95 I -> L (IN ISOLATE R1047).  
FT VARIANT 135 135 E -> K (IN ISOLATE R1042).  
FT VARIANT 142 142 E -> H (IN ISOLATE R1042).  
FT VARIANT 154 154 F -> S (IN ISOLATE R1042).  
SO SEQUENCE 160 AA: 18912 MW: 144997E CRC32;  
Query Match 45.1%; Score 41; DB 1; Length 160;  
Best Local Similarity 63.6%; Pred. No. 4.9;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
OY 3 GRKVELGEPYC 13  
DB 44 GRKTESIGRP 54  
RESULT 14  
YB6A\_SCHPO  
ID YB6A\_SCHPO STANDARD: PRT: 1522 AA.  
AC Q09750:  
DT 01-NOV-1995 (REL. 32, CREATED)  
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE HYPOTHEICAL 171.5 KD PROTEIN C12C2.10C IN CHROMOSOME II.  
GN SPIC12C2.10C.  
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
OC EUKARYOTA: FUNGI: ASCOMYCOTINA: HEMIASCOMYCETES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-972;

RA DEVLIN K., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;  
RL SUBMITTED (SEP-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- SIMILARITY: TO YEAST PAIRED AMPHIPATHIC HELIX PROTEIN (SIN3).  
DR EMBL: 254140: G984217: -.  
KW HYPOTHEICAL PROTEIN.  
SO SEQUENCE 1522 AA: 171455 MW: 1D563B22 CRC32;  
Query Match 45.1%; Score 41; DB 1; Length 1522;  
Best Local Similarity 50.0%; Pred. No. 49;  
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
OY 4 KRVELGEPYC 17  
DB 999 RSVYNTLFGNMSLYC 1012  
RESULT 15  
CR2\_MOUSE  
ID CR2\_MOUSE STANDARD: PRT: 1025 AA.  
AC P19070:  
DT 01-NOV-1990 (REL. 16, CREATED)  
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)  
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
DE COMPLEMENT RECEPTOR TYPE 2 PRECURSOR (CR2) (COMPLEMENT C3D RECEPTOR).  
GN CR2.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA;  
OC EUTHERIA: RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BALB/C;  
RX MEDLINE: 90229735.  
RA FINGEROTH J.D.;  
RL J. IMMUNOL. 144:3458-3467(1990).  
RN [2]  
RP SEQUENCE OF 12-1025 FROM N.A.  
RX MEDLINE: 91010789.  
RA MOLINA H., KINOSHITA T., INOUE K., CAREL J.C., HOLERS V.M.;  
RL J. IMMUNOL. 145:2974-2983(1990).  
RN [3]  
RP SEQUENCE OF 343-401 AND 991-1025 FROM N.A.  
RX MEDLINE: 89098890.  
RA FINGEROTH J.D., BENEDICT M.A., LEVY D.N., STROMINGER J.L.;  
RL PROC. NATL. ACAD. SCI. U.S.A. 86:242-246(1989).  
RN [4]  
RP SEQUENCE OF 289-1025 FROM N.A.  
RX MEDLINE: 89381350.  
RA KURTZ C.B., PAUL M.S., AEGERTER M., WEIS J.J., WEIS J.H.;  
RL J. IMMUNOL. 143:2058-2067(1989).  
CC -1- FUNCTION: RECEPTOR FOR COMPLEMENT C3D. PARTICIPATES IN B  
CC LYMPHOCYTES ACTIVATION.  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -1- TISSUE SPECIFICITY: B LYMPHOCYTES.  
CC -1- SIMILARITY: TO HUMAN C3D/EBV RECEPTOR (CD21).  
DR EMBL: M81083: G192697: -.  
DR EMBL: M35684: G192688: -.  
DR EMBL: M61132: G192693: -.  
DR EMBL: M35685: G192695: ALT\_SEQ.  
DR EMBL: M29281: G387131: -.  
DR PIR: A43526: A43526.  
DR HSSP: P08603: 1HCC.  
DR MGD: MGT:88489: CR2.  
KW COMPLEMENT PATHWAY; GLYCOPROTEIN; TRANSMEMBRANE; REPEAT; SIGNAL;  
KW RECEPTOR; SUSHI.  
FT SIGNAL 1 11  
FT CHAIN 12 1025  
FT DOMAIN 12 963  
FT TRANSMEM 964 990  
FT DOMAIN 991 1025  
FT DOMAIN 13 1025  
FT DOMAIN 13 959  
FT REPEAT 13 74  
SUSHI 1. COMPLEMENT RECEPTOR TYPE 2.  
EXTRACELLULAR (POTENTIAL).  
POTENTIAL.  
CYTOPLASMIC (POTENTIAL).  
15 X SUSHI (SCR). REPEATS.



FT REPEAT 81 139 SUSHI 2.  
FT REPEAT 145 203 SUSHI 3.  
FT REPEAT 206 264 SUSHI 4.  
FT REPEAT 267 335 SUSHI 5.  
FT REPEAT 342 399 SUSHI 6.  
FT REPEAT 401 458 SUSHI 7.  
FT REPEAT 462 515 SUSHI 8.  
FT REPEAT 518 586 SUSHI 9.  
FT REPEAT 593 649 SUSHI 10.  
FT REPEAT 653 705 SUSHI 11.  
FT REPEAT 708 769 SUSHI 12.  
FT REPEAT 777 834 SUSHI 13.  
FT REPEAT 840 898 SUSHI 14.  
FT REPEAT 901 959 SUSHI 15.  
FT DISULFID 14 56 BY SIMILARITY.  
FT DISULFID 42 73 BY SIMILARITY.  
FT DISULFID 82 124 BY SIMILARITY.  
FT DISULFID 110 138 BY SIMILARITY.  
FT DISULFID 146 189 BY SIMILARITY.  
FT DISULFID 175 202 BY SIMILARITY.  
FT DISULFID 207 248 BY SIMILARITY.  
FT DISULFID 234 263 BY SIMILARITY.  
FT DISULFID 268 317 BY SIMILARITY.  
FT DISULFID 297 334 BY SIMILARITY.  
FT DISULFID 343 385 BY SIMILARITY.  
FT DISULFID 371 398 BY SIMILARITY.  
FT DISULFID 402 445 BY SIMILARITY.  
FT DISULFID 431 458 BY SIMILARITY.  
FT DISULFID 463 501 BY SIMILARITY.  
FT DISULFID 487 514 BY SIMILARITY.  
FT DISULFID 519 568 BY SIMILARITY.  
FT DISULFID 548 585 BY SIMILARITY.  
FT DISULFID 594 636 BY SIMILARITY.  
FT DISULFID 622 649 BY SIMILARITY.  
FT DISULFID 654 689 BY SIMILARITY.  
FT DISULFID 675 704 BY SIMILARITY.  
FT DISULFID 709 752 BY SIMILARITY.  
FT DISULFID 738 769 BY SIMILARITY.  
FT DISULFID 778 820 BY SIMILARITY.  
FT DISULFID 806 833 BY SIMILARITY.  
FT DISULFID 841 884 BY SIMILARITY.  
FT DISULFID 870 897 BY SIMILARITY.  
FT DISULFID 902 945 BY SIMILARITY.  
FT DISULFID 931 958 BY SIMILARITY.  
FT CARBOHYD 77 77 POTENTIAL.  
FT CARBOHYD 113 113 POTENTIAL.  
FT CARBOHYD 276 276 POTENTIAL.  
FT CARBOHYD 316 316 POTENTIAL.  
FT CARBOHYD 364 364 POTENTIAL.  
FT CARBOHYD 380 380 POTENTIAL.  
FT CARBOHYD 484 484 POTENTIAL.  
FT CARBOHYD 527 527 POTENTIAL.  
FT CARBOHYD 615 615 POTENTIAL.  
FT CARBOHYD 639 639 POTENTIAL.  
FT CARBOHYD 694 694 POTENTIAL.  
FT CARBOHYD 754 754 POTENTIAL.  
FT CARBOHYD 790 790 POTENTIAL.  
FT CARBOHYD 813 813 POTENTIAL.  
FT CARBOHYD 823 823 POTENTIAL.  
FT CARBOHYD 851 851 POTENTIAL.  
FT CARBOHYD 901 901 POTENTIAL.  
FT CONFLICT 289 291 YGS -> EPR (IN REF. 4).  
FT CONFLICT 306 306 S -> T (IN REF. 2).  
FT CONFLICT 520 520 P -> A (IN REF. 2).  
FT CONFLICT 962 963 MISSING (IN REF. 4).  
SQ SEQUENCE 1025 AA; 112994 MW; DA043E62 CRC32;

Db 558 FKLGEOIHC 568

Search completed: July 18, 1999, 11:37:52  
Job time: 1751 sec

Query Match 44.0%; Score 40; DB 1; Length 1025;  
Best Local Similarity 54.5%; Pred. No. 49;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
Oy 7 FELVGEPSIYC 17



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd

Run on: July 18, 1999, 06:02:50 ; Search time 63.5 Seconds

14.770 Million cell updates/sec

Title: US-09-142-043-5

Sequence: 1 SGGRKVFELVGEPSTYC 17

Scoring table: BLOSUM62

Searched: 180763 seqs, 55169189 residues

Database

```

1: SPTRMBL.8:*
2: sp_fungi:*
3: sp_human:*
4: sp_invertebrate:*
5: sp_mammal:*
6: sp_mhc:*
7: sp_ornamente:*
8: sp_phase:*
9: sp_plant:*
10: sp_bacteria:*
11: sp_rhodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_archaea:*

```

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	91	100.0	2039	2	016745	016745 homo sapiens
2	91	100.0	2489	2	016744	016744 homo sapiens
3	91	100.0	461	4	029531	029531 pan troglodytes
4	91	100.0	2014	4	029530	029530 pan troglodytes
5	88	96.7	315	4	028770	028770 papio hama
6	88	96.7	1911	4	029528	029528 papio hama
7	88	96.7	522	4	028797	028797 pan troglodytes
8	80	87.9	479	2	014079	014079 homo sapiens
9	77	84.6	522	4	028769	028769 papio hama
10	65	71.4	417	10	035520	035520 rattus norvegicus
11	65	71.4	89	10	063129	063129 rattus norvegicus
12	65	71.4	497	10	063612	063612 rattus norvegicus
13	65	71.4	559	10	063135	063135 rattus norvegicus
14	62	68.1	483	10	064735	064735 mus musculus
15	56	61.5	679	10	099254	099254 mus musculus
16	46	50.5	222	4	019120	019120 actus trivialis
17	46	50.5	222	4	019125	019125 salmella salmella
18	45	49.5	349	2	015429	015429 homo sapiens
19	45	49.5	347	2	0178361	0178361 homo sapiens
20	45	49.5	734	3	017237	017237 caenorhabditis elegans
21	45	49.5	336	4	062834	062834 sagittus ocellatus
22	45	49.5	314	4	062835	062835 sagittus ocellatus
23	45	49.5	378	4	062837	062837 sagittus ocellatus
24	45	49.5	377	4	062838	062838 sagittus ocellatus
25	45	49.5	363	4	002839	002839 sus scrofa
26	45	49.5	222	4	019122	019122 callinectes sapidus
27	45	49.5	222	4	019123	019123 callinectes sapidus
28	45	49.5	222	4	019124	019124 sagittus ocellatus
29	45	49.5	222	4	019128	019128 pithecia pithecia

	30	4	48.4	611	9	066634	066634 aquifex ae
	31	44	48.4	1441	11	0833001	0833001 bunyavirus
	32	43	47.3	285	4	019121	019121 papio hamad
	33	43	47.3	285	4	019126	019126 macaca fasc
	34	43	47.3	285	4	019127	019127 macaca mela
	35	43	47.3	369	4	P79138	P79138 ceropithhec
	36	42	46.2	384	3	061847	061847 caenorhabdi
	37	42	46.2	165	9	Q05170	Q05170 escherichia
	38	42	46.2	390	10	P97732	P97732 mus musculu
	39	41	45.1	160	7	O64368	O64368 lactobacill
	40	41	45.1	2611	9	O68487	O68487 streptomyce
	41	41	45.1	724	9	O83823	O83823 treponema p
	42	41	45.1	666	14	O27798	O27798 methanobact
	43	40	44.0	556	8	O82722	O82722 nicotiana s
	44	39.5	43.4	435	2	O75901	O75901 homo sapien
	45	39.5	43.4	341	9	O32456	O32456 actinomadar
 ALIGNMENTS							
RESULT	1						
ID	016745	PRELIMINARY;	PRT; 2039 AA.				
AC	016745;						
DT	01-NOV-1996 (TREMBLREL, 01, CREATED)						
DT	01-NOV-1996 (TREMBLREL, 01, LAST SEQUENCE UPDATE)						
DT	01-NOV-1998 (TREMBLREL, 08, LAST ANNOTATION UPDATE)						
DE	COMPLEMENT RECEPTOR 1.						
OS	HOMO SAPIENS (HUMAN).						
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;						
CC	CARABRHINI; HOMINIDAE; HOMO.						
RN	[1]						
RA	SEQUENCE FROM N.A.						
RX	MEDLINE; 94065175.						
SA	VIK D.P., WONG W.W.;						
RT	"Structure of the gene for the F allele of complement receptor type 1						
RL	J. IMMUNOL. 151:6214-6224(1993).						
RN	[2]						
RP	SEQUENCE FROM N.A.						
RA	VIK D.P., WONG W.W.;						
RL	SUBMITTED (JUN-1993) TO EMBL/GENBANK/DDBJ DATA BANKS.						
DR	EMBL; L17399; G306680; JOINED.						
DR	EMBL; L17409; G306680; JOINED.						
DR	EMBL; L17419; G306680; JOINED.						
DR	EMBL; L17420; G306680; JOINED.						
DR	EMBL; L17421; G306680; JOINED.						
DR	EMBL; L17422; G306680; JOINED.						
DR	EMBL; L17423; G306680; JOINED.						
DR	EMBL; L17391; G306680; JOINED.						
DR	EMBL; L17392; G306680; JOINED.						
DR	EMBL; L17393; G306680; JOINED.						
DR	EMBL; L17394; G306680; JOINED.						
DR	EMBL; L17395; G306680; JOINED.						
DR	EMBL; L17396; G306680; JOINED.						
DR	EMBL; L17397; G306680; JOINED.						
DR	EMBL; L17398; G306680; JOINED.						
DR	EMBL; L17418; G306680; JOINED.						
DR	EMBL; L17390; G306680; JOINED.						
DR	EMBL; L17400; G306680; JOINED.						
DR	EMBL; L17416; G306680; JOINED.						
DR	EMBL; L17417; G306680; JOINED.						
DR	EMBL; L17401; G306680; JOINED.						
DR	EMBL; L17402; G306680; JOINED.						
DR	EMBL; L17403; G306680; JOINED.						
DR	EMBL; L17404; G306680; JOINED.						
DR	EMBL; L17405; G306680; JOINED.						
DR	EMBL; L17406; G306680; JOINED.						
DR	EMBL; L17407; G306680; JOINED.						
DR	EMBL; L17408; G306680; JOINED.						
DR	EMBL; L17410; G306680; JOINED.						
DR	EMBL; L17411; G306680; JOINED.						

DR EMBL; L17412; G306680; JOINED.  
DR EMBL; L17413; G306680; JOINED.  
DR EMBL; L17414; G306680; JOINED.  
DR EMBL; L17415; G306680; JOINED.  
DR PFAM; PF00084; sushi; 30.  
SQ SEQUENCE 2039 AA; 223603 MW; A4972215 CRC32;

Query Match  
Best Local Similarity 100.0%; Score 91; DB 2; Length 2039;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGRKRVFELVGEPSIYC 17  
|||||  
Db 199 SGRKRVFELVGEPSIYC 215

RESULT 2  
ID Q16744 PRELIMINARY; PRT; 2489 AA.

AC Q16744;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE COMPLEMENT RECEPTOR 1.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; HOMO.  
RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE; 94065175.  
RA VIR D.P., WONG W.W.;  
RT "Structure of the gene for the F allele of complement receptor type 1  
and structure of the coding region unique to the S allele."  
J. IMMUNOL. 151:6214-6224(1993).

RL [2]  
RN SEQUENCE FROM N.A.  
RA VIR D.P., WONG W.W.;  
RL SUBMITTED (JUN-1993) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL; L17399; G451303; JOINED.  
DR EMBL; L17409; G451303; JOINED.  
DR EMBL; L17419; G451303; JOINED.  
DR EMBL; L17420; G451303; JOINED.  
DR EMBL; L17421; G451303; JOINED.  
DR EMBL; L17422; G451303; JOINED.  
DR EMBL; L17423; G451303; JOINED.  
DR EMBL; L17424; G451303; JOINED.  
DR EMBL; L17425; G451303; JOINED.  
DR EMBL; L17426; G451303; JOINED.  
DR EMBL; L17427; G451303; JOINED.  
DR EMBL; L17428; G451303; JOINED.  
DR EMBL; L17429; G451303; JOINED.  
DR EMBL; L17430; G451303; JOINED.  
DR EMBL; L17391; G451303; JOINED.  
DR EMBL; L17392; G451303; JOINED.  
DR EMBL; L17393; G451303; JOINED.  
DR EMBL; L17394; G451303; JOINED.  
DR EMBL; L17395; G451303; JOINED.  
DR EMBL; L17396; G451303; JOINED.  
DR EMBL; L17397; G451303; JOINED.  
DR EMBL; L17398; G451303; JOINED.  
DR EMBL; L17400; G451303; JOINED.  
DR EMBL; L17401; G451303; JOINED.  
DR EMBL; L17402; G451303; JOINED.  
DR EMBL; L17403; G451303; JOINED.  
DR EMBL; L17405; G451303; JOINED.  
DR EMBL; L17406; G451303; JOINED.  
DR EMBL; L17407; G451303; JOINED.  
DR EMBL; L17408; G451303; JOINED.

DR EMBL; L17410; G451303; JOINED.  
DR EMBL; L17411; G451303; JOINED.  
DR EMBL; L17412; G451303; JOINED.  
DR EMBL; L17413; G451303; JOINED.  
DR EMBL; L17414; G451303; JOINED.  
DR EMBL; L17415; G451303; JOINED.  
DR PFAM; PF00084; sushi; 37.  
SQ SEQUENCE 2489 AA; 272846 MW; 5869B6F9 CRC32;

Query Match  
Best Local Similarity 100.0%; Score 91; DB 2; Length 2489;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGRKRVFELVGEPSIYC 17  
|||||  
Db 199 SGRKRVFELVGEPSIYC 215

RESULT 3  
ID Q29531 PRELIMINARY; PRT; 661 AA.

AC Q29531;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE COMPLEMENT RECEPTOR 1 PRECURSOR (FRAGMENT).  
GN CRI.  
OS PAN TROGLODYTES (CHIMPANZEE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; PAN.  
RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE; 94292799.  
RA BIRMINGHAM D.J., SHEN X.P., HOURCADE D., NICKELLS M.W., ATKINSON J.P.;  
RT "Primary sequence of an alternatively spliced form of CRI. Candidate  
for the 75,000 M(r) complement receptor expressed on chimpanzee  
erythrocytes."  
J. IMMUNOL. 153:691-700(1994).

DR EMBL; L24921; G557727; -  
DR PFAM; PF00084; sushi; 9.  
KW SIGNAL; ALTERNATIVE SPLICING.  
FT NON\_TER 1  
FT SIGNAL 1  
FT SIGNAL 16  
FT CHAIN 17  
FT NON\_TER 661  
SQ SEQUENCE 661 AA; 72966 MW; 9D78E262 CRC32;

Query Match  
Best Local Similarity 100.0%; Score 91; DB 4; Length 661;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGRKRVFELVGEPSIYC 17  
|||||  
Db 174 SGRKRVFELVGEPSIYC 190

RESULT 4  
ID Q29530 PRELIMINARY; PRT; 2014 AA.

AC Q29530;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE COMPLEMENT RECEPTOR 1 (FRAGMENT).  
GN CRI.  
OS PAN TROGLODYTES (CHIMPANZEE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; PAN.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 94292799.

RA BIRMINGHAM D.J., SHEN X.P., HOURCADE D., NICKELLS M.W., ATKINSON J.P.;  
RT "Primary sequence of an alternatively spliced form of CRL. Candidate  
for the 75,000 M(r) complement receptor expressed on chimpanzee  
erythrocytes."  
RT J. IMMUNOL. 153:691-700(1994).  
DR EMBL: L24920; G55725; -  
DR PFAM: PF00084; sushi; 30.  
FT NON\_TER 1  
FT NON\_TER 2014  
SQ SEQUENCE 2014 AA; 221280 MW; 1370CDD8 CRC32;

Query Match 100.0%; Score 91; DB 4; Length 2014;  
Best Local Similarity 100.0%; Pred. No. 2.8e-07;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SGGKRVFELVGPSTYC 17  
Db 174 SGGKRVFELVGPSTYC 190

RESULT 5  
O28770 PRELIMINARY; PRT; 315 AA.  
AC O28770;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE COMPLEMENT RECEPTOR (FRAGMENT).  
OS PAPIO HAMADRYAS (HAMADRYAS BABOON).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; PAPIO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BONE MARROW;  
RA BIRMINGHAM D.J., LOGAR C.M., SHEN X.P., CHEN W.;  
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: L77978; G1301611; -  
DR PFAM: PF00084; sushi; 5.  
FT NON\_TER 1  
FT NON\_TER 315  
SQ SEQUENCE 315 AA; 35172 MW; 5776DB8A CRC32;

Query Match 96.7%; Score 88; DB 4; Length 315;  
Best Local Similarity 94.1%; Pred. No. 1.4e-07;  
Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SGGKRVFELVGPSTYC 17  
Db 158 SGGKRVFELVGPSTYC 174

RESULT 6  
O29528 PRELIMINARY; PRT; 1911 AA.  
AC O29528;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE COMPLEMENT RECEPTOR 1 (FRAGMENT).  
OS CRL.  
GN PAPIO HAMADRYAS (HAMADRYAS BABOON).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; PAPIO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA CLEMENZA L., SUBRAMANIAN B.V., NICKELLS M.W., HOURCADE D.E.,  
RA ATKINSON J.P.;  
RL SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: L39791; G662829; -  
DR PFAM: PF00084; sushi; 29.  
FT NON\_TER 1911

SQ SEQUENCE 1911 AA; 210173 MW; 03D640C2 CRC32;

Query Match 96.7%; Score 88; DB 4; Length 1911;  
Best Local Similarity 94.1%; Pred. No. 8.9e-07;  
Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SGGKRVFELVGPSTYC 17  
Db 86 SGGKRVFELVGPSTYC 102

RESULT 7  
O28797 PRELIMINARY; PRT; 522 AA.  
AC O28797;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE UNKNOWN PROTEIN (FRAGMENT).  
OS PAN TROGLODYTES (CHIMPANZEE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; PAN.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 94292799.  
RA BIRMINGHAM D.J., SHEN X.P., HOURCADE D., NICKELLS M.W., ATKINSON J.P.;  
RT "Primary sequence of an alternatively spliced form of CRL. Candidate  
for the 75,000 M(r) complement receptor expressed on chimpanzee  
erythrocytes."  
RT J. IMMUNOL. 153:691-700(1994).  
RL EMBL: L24922; G557729; -  
DR PFAM: PF00084; sushi; 8.  
FT NON\_TER 1  
SQ SEQUENCE 522 AA; 57304 MW; 21A32E12 CRC32;

Query Match 96.7%; Score 88; DB 4; Length 522;  
Best Local Similarity 94.1%; Pred. No. 2.3e-07;  
Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SGGKRVFELVGPSTYC 17  
Db 141 SGGKRVFELVGPSTYC 157

RESULT 8  
O14079 PRELIMINARY; PRT; 479 AA.  
AC O14079;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE COMPLEMENT RECEPTOR 1-LIKE PROTEIN (CRL) (FRAGMENT).  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE: 90110163.  
RA HOURCADE D., MIESNER D.R., BEE C., ZELDES W., ATKINSON J.P.;  
RT "Duplication and divergence of the amino-terminal coding region of the  
complement receptor 1 (CRL) gene. An example of concerted (horizontal)  
evolution within a gene."  
RT J. BIOL. CHEM. 265:974-980(1990).  
RL EMBL: M31231; G563326; JOINED.  
DR EMBL: M31232; G563326; JOINED.  
DR EMBL: M31233; G563326; JOINED.  
DR EMBL: M31237; G563326; -  
DR EMBL: M31230; G563326; JOINED.  
DR EMBL: M31234; G563326; JOINED.  
DR EMBL: X14359; G563326; JOINED.  
DR EMBL: X14360; G563326; JOINED.





KW ALTERNATIVE SPLICING; RECEPTOR.  
FT SIGNAL 1 11  
FT CHAIN 12 679 COMPLEMENT RECEPTOR TYPE 2.  
FT NON TER 679 679  
SQ SEQUENCE 679 AA; 74916 MW; BF08AC75 CRC32;

Query Match 61.5%; Score 56; DB 10; Length 679;  
Best Local Similarity 76.9%; Pred. No. 0.11;  
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 KVEELVGEPSIYC 17  
I:I:|||||  
Db 182 KLPDLVGKSIYC 194

Search completed: July 18, 1999, 06:02:50  
Job time: 3105 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 00:42:54 ; Search time 58.51 Seconds  
(without alignments)  
2.420 Million cell updates/sec

Title: US-09-142-043-6

Perfect score: 34  
Sequence: 1 KKKKKKA 7

Scoring table: BLOSUM62

Searched: 162890 seqs, 20225328 residues

Database : A\_Geneseq\_34:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	34	100.0	16	1	R65867	Antideoxyribonucle
2	34	100.0	17	1	R66010	Peptide which bind
3	31	91.2	16	1	P60118	K4B-ras p21 oncoge
4	31	91.2	481	1	R33389	AlI/AVPv2 receptor
5	31	91.2	141	1	R66394	Human psychostis pr
6	31	91.2	24	1	R95575	Protein K1-ras 2B
7	31	91.2	576	1	R85891	WD-40 domain-contg
8	31	91.2	228	1	R95675	K-ras oncoprotein.
9	31	91.2	116	1	W03642	Human cannabinoide
10	31	91.2	14	1	W04466	Prenyl acceptor pe
11	31	91.2	15	1	W04468	Human K-ras B prot
12	31	91.2	188	1	W04473	Human K-ras B prot
13	31	91.2	1040	1	W24559	Presentin-Interac
14	31	91.2	17	1	W34654	C-terminal 17 amin
15	31	91.2	214	1	W45898	SCR 1-3 of complm
16	31	91.2	17	1	W45890	Peptide membrane b
17	31	91.2	17	1	W45892	Peptide membrane b
18	31	91.2	15	1	W45893	Peptide membrane b
19	31	91.2	16	1	W45878	Peptide membrane b
20	31	91.2	16	1	W45881	Peptide membrane b
21	31	91.2	14	1	W45882	Peptide membrane b
22	31	91.2	18	1	W58548	K-Ras4B farnesylat
23	31	91.2	756	1	W60664	Human ALARM protei
24	31	91.2	215	1	W75987	Complement recepto
25	31	91.2	215	1	W75990	(CM15)-Cys-S-S-(MS
26	31	91.2	215	1	W75992	(CM16)-Cys-S-S-(MS
27	30	88.2	783	1	P94265	Sequence of APh36.
28	30	88.2	694	1	R04107	DNA-binding protei
29	30	88.2	744	1	P81006	Alpha-2-plasmin in
30	30	88.2	175	1	R06261	Peptide antigenic
31	30	88.2	12	1	R24220	Fragment of tenasc
32	30	88.2	12	1	R24217	Fragment of tenasc
33	30	88.2	12	1	P71160	Peptide which inh
34	30	88.2	12	1	P71165	Peptide which inh
35	30	88.2	29	1	R11512	Anti-ATLA antibody
36	30	88.2	29	1	R11513	Anti-ATLA antibody
37	30	88.2	34	1	R11514	Anti-ATLA antibody
38	30	88.2	898	1	P61030	Entire coded sequ
39	30	88.2	899	1	P61056	Translation of pla
40	30	88.2	898	1	P61082	Complete translat
41	30	88.2	225	1	P60453	Sequence of the Ma
42	30	88.2	317	1	R21750	C-terminal sequenc
43	30	88.2	242	1	R22390	Antigen ac-1b. Vac

## ALIGNMENTS

RESULT 1  
R65867 standard; peptide: 16 AA.  
ID R65867:  
AC R65867:  
DT 26-JUN-1995 (first entry)  
DE Antideoxyribonucleic acid antibody adsorbent peptide.  
KW Antideoxyribonucleic acid adsorbent peptide; kidney inflammation;  
KW systemic lupus erythematosus.  
OS Synthetic.  
PN J06261941-A.  
PD 20-SEP-1994.  
PF 13-OCT-1993; 256111.  
PR 18-JAN-1993; JP-006100.  
PA (KURS ) KURARAY CO LTD.  
DR WPI: 94-337461/42.  
PT Adsorbents of antideoxyribonucleic - comprising a peptide,  
PT immobilised on a carrier.  
PS Claim 1; Page 15; 16pp; Japanese.  
CC R65841-R65879 are antideoxyribonucleic (ADN) acid adsorbent peptides  
CC derived from the highly generic peptide in the specification. When these  
CC peptides are immobilised on a carrier (either cellulose, porous  
CC glass, polyvinyl alcohol or polyacrylamide) they can be used to  
CC remove ADN antibodies, which can result in kidney inflammation and  
CC systemic lupus erythematosus.  
SQ Sequence 16 AA:

Query Match 100.0%; Score 34; DB 1; Length 16;  
Best Local Similarity 100.0%; Pred. No. 2.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 KKKKKKA 7  
DB 5 KKKKKKA 11

RESULT 2  
R66010 standard; peptide: 17 AA.  
ID R66010:  
AC R66010:  
DT 30-JUN-1995 (first entry)  
DE Peptide which binds immune complexes.  
KW Immune complex; immunoadsorbent.  
OS Synthetic.  
PN J06263795-A.  
PD 20-SEP-1994.  
PF 30-SEP-1993; 245367.  
PR 30-SEP-1992; JP-261821.  
PR 18-JAN-1993; JP-006099.  
PA (KURS ) KURARAY CO LTD.  
DR WPI: 94-338300/42.  
PT Peptide(s) capable of binding to immune complex, and adsorbents  
PT contg them - useful for removal of immune complexes from body  
PT fluid and for treatment of associated diseases.  
PS Example 12; Page 6; 15pp; Japanese.  
CC A is a peptide fragment of 6-12 amino acids containing the sequence  
CC Ala-B-C-Glu-Ile-Leu, where B and C are Trp, Tyr and/or Phe; X and Y  
CC are each a single bond, or an amino acid or peptide composed of 2-10  
CC amino acids selected from Asp, Glu, Arg, Lys and His, provided at  
CC least one of X and Y is not a bond; and Z is OH or amide.  
CC The peptides can bind to immune complexes. They can be fixed to a  
CC support and used for selective removal of immune complexes from body  
CC fluids without removal of useful components. They can thus be used for  
CC treating diseases involving the complexes. In Examples, the  
CC activity of the peptides is compared to that of reference peptides



CC in which expression is greatest in the 'well' twin. A nt  
 CC sequence substantially corresp. to Q75090 is claimed, which  
 CC comprises 30-X nts, where X=80, 95, 158, 222, 249, 260, 295,  
 CC 407 or 423. A PP peptide of 10-141 AAs is also claimed.  
 SO Sequence 141 AA;

Query Match 91.2%; Score 31; DB 1; Length 141;  
 Best Local Similarity 85.7%; Pred. No. 46;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKA 7  
 Db 5 KKKKKKS 11

RESULT 6  
 ID R95575  
 AC R95575; standard; peptide; 24 AA.  
 DT 06-SEP-1996 (first entry)  
 DE Protein Ki-ras 2B hypervariable region epitope.  
 KM Intracellular binding protein; fusion protein; heavy chain; light chain;  
 KM variable region; ras; monoclonal antibody; linker; Pelt leader sequence;  
 KM PCR; polymerase chain reaction; amplification; expression vector; cancer;  
 KM phagemid; mammalian cell; structural formation; epitope; vaccination;  
 KM gene therapy; viral infection; virus.  
 OS Synthetic.  
 PN MO9429446-A2.  
 PD 22-DEC-1994.  
 PF 15-JUN-1994; E00714.  
 PR 16-JUN-1993; FR-007241.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Schweighoffer F, Tocque B;  
 DR WPI: 95-036475/05.  
 PT Nucleic acid encoding intracellular binding protein - partic.  
 PT antibody or deriv., under control of promoter functional in  
 PT mammalian cells, for gene therapy of cancer and virus disease  
 PS Example 3; Page 20; 32pp; French.  
 CC A novel intracellular binding protein of the invention is generated by  
 CC fusing the heavy and light chains variable regions of an antibody  
 CC e.g. The anti-ras monoclonal antibody Y13-259 (R95569) or an anti-GAP  
 CC antibody (R95570). The peptides R95574-5 are synthetic peptides derived  
 CC from the Ki-ras proteins 2A or 2B hypervariable regions, used to generate  
 CC antibodies against those proteins. The sequences encoding the Vh and Vk  
 CC regions are then used to construct fusion antibodies. The fusion proteins  
 CC were synthesised by PCR amplifying the Vh and Vk coding sequences and  
 CC ligating them with a linker sequence into an expression vector such as a  
 CC phagemid e.g. M13. The sequences can also be inserted into a mammalian  
 CC cell for expression of the protein. The novel proteins are expressed  
 CC intracellularly and can target epitopes not accessible by the usual  
 CC vaccination methods. The protein can thus be used for gene therapy e.g.  
 CC of cancer or viral infections.  
 SO Sequence 24 AA;

Query Match 91.2%; Score 31; DB 1; Length 24;  
 Best Local Similarity 85.7%; Pred. No. 10;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKA 7  
 Db 11 KKKKKKS 17

RESULT 7  
 ID R85891  
 AC R85891; standard; Protein; 576 AA.  
 DT 13-SEP-1996 (first entry)  
 DE MD-40 domain-contg. periodic TTP protein.  
 KM WD40 repeat region; beta-transducin; protein-protein interaction; drug;  
 KM intracellular signaling; protein kinase C; homology; motif; modulator;

KM receptors of activated protein kinase; enzyme activity; isozyme; human.  
 OS Synthetic.  
 PN MO9521252-A2.  
 PD 10-AUG-1995.  
 PF 31-JAN-1995; U01210.  
 PR 01-FEB-1994; US-190802.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Mochly-Rosen D, Ron D;  
 DR WPI: 95-283772/37.  
 PT New WD-40 (beta-transducin)-derived polypeptide(s) - which alter the  
 PT activity of a protein, eg. protein kinase C, which interacts with a  
 PT protein contg. a WD-40 region.  
 PS Example 5; Page 141-143; 351pp; English.  
 CC Proteins R85851-92 are protein which contain at least one WD-40 (also  
 CC called beta-transducin homologous) amino acid repeat motifs. The WD-40  
 CC regions are involved in protein-protein interactions between proteins  
 CC involved in intracellular signaling. An example of such an interaction  
 CC is between protein kinase C and receptors of activated protein kinase  
 CC (RACK), esp. RACK-1 (R85850). Proteins R85851-82 were isolated based on  
 CC homology with beta-transducin, whereas proteins R85882-92 were isolated  
 CC based on homology with the WD-40 consensus sequence (R85893). The  
 CC proteins were used to construct the peptides R84928-R85063 and  
 CC R85786-R85842. The peptides can be used to identify target proteins  
 CC contg. WD-40 motifs, as modulators of enzyme esp. isozyme, activity of  
 CC proteins involved in protein-protein interaction and to screen for drugs  
 CC that will affect protein-protein interaction involving WD-40 domains.  
 SO Sequence 576 AA;

Query Match 91.2%; Score 31; DB 1; Length 576;  
 Best Local Similarity 85.7%; Pred. No. 1; Seq=02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKA 7  
 Db 270 KKKKKKS 276

RESULT 8  
 ID R95675  
 AC R95675; standard; Protein; 228 AA.  
 DT 26-OCT-1996 (first entry)  
 DE K-ras oncogene.  
 KM Human; K-ras; oncogene; activating mutation; restriction site;  
 KM PCR; polymerase chain reaction; DNA; 35R; Q-beta-replicase;  
 KM colorectal cancer; diagnostic; prognosis.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT misc\_difference 190  
 FT "in-frame stop codon"  
 PN MO9615262-A2.  
 PD 23-MAY-1996.  
 PF 10-NOV-1995; G02644.  
 PR 11-NOV-1994; GB-022814.  
 PA (DZIE/) DZIEGLEWSKA H E.  
 PI (MED1-) MEDINNOVA SF.  
 PI Breivik J, Gaudernack G;  
 DR WPI: 96-259858/26.  
 DR N-PSDB: T31723.  
 PT Primers to detect activating mutation(s) of the K-ras oncogene  
 PT useful to screen for cancers, partic. colorectal cancer  
 PS Example 1; Fig 1; 39pp; English.  
 CC This sequence represents a human cellular wild-type K-ras  
 CC oncoprotein, encoded by a protooncogene c-k-ras2 sequence.  
 CC The sequence has been used to construct primers T31717-22  
 CC (claimed), which have been designed with specific mismatches  
 CC to introduce restriction sites overlapping upstream and/or  
 CC downstream of activating mutation hotspot sites at codons 12,  
 CC 13 and/or 61 during in vitro amplification (PCR, DNA, 35R  
 CC and Q-beta-replicase methods claimed). The primers are useful  
 CC in screening for e.g. colorectal cancer.  
 SO Sequence 228 AA;

Query Match 91.2%; Score 31; DB 1; Length 228;  
Best Local Similarity 85.7%; Pred. No. 69;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKRKA 7  
|||||:  
DB 215 KKKRKS 221

RESULT 9  
W03642  
ID W03642 standard; peptide; 116 AA.  
AC W03642;  
DT 17-DEC-1996 (first entry)  
DE Human cannabinoid GPR N-terminal sequence.  
KW G-protein coupled receptor; ligand binding assay; transmembrane domain;  
KW schizophrenia; dopamine; CAMP; adenosine; thrombin; adrenergic; opsin;  
KW muscarinic acetylcholine; endothelin; bombesin; endocrine; rhodopsin;  
KW odorant; cytomegalovirus; serotonergic.  
OS Homo sapiens.  
PN US508384-A.  
PD 16-APR-1996.  
PF 10-SEP-1992; 943236.  
PR 10-SEP-1992; US-943236.  
PR 09-SEP-1993; US-118270.  
PA (UWNY ) UNIV NEW YORK STATE.  
PI Murphy RB, Schuster DJ;  
DR WPI: 96-208785/21.  
PT New dopamine receptor peptide - useful as antipsychotic agent, e.g.  
PT for treating schizophrenia  
PS Disclosure: Fig 8B(3): 184pp; English.  
CC Proteins W02657-W02720 represent a range of G-protein coupled receptor  
CC (GPR) proteins selected from CAMP, adenosine, muscarinic acetylcholine,  
CC adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,  
CC odorant, cytomegaloviral and other GPR proteins. The peptides  
CC W03578-W03651 represent the N-terminal fragments of the above proteins.  
CC The receptor proteins were used to design polypeptides, pref. based on  
CC the transmembrane domains, for use in G-protein coupled receptor ligand  
CC binding assays. The polypeptide fragments retain biological activity  
CC such as binding a GPR ligand or modulating GPR ligand binding to a GPR  
CC (see W02747-W02999 for examples of polypeptide fragments).  
CC The polypeptide fragments can be used in compositions for treating  
CC subjects suffering from a pathology related to a GPR abnormality e.g. a  
CC psychotic disorder such as schizophrenia.  
SQ Sequence 116 AA;

Query Match 91.2%; Score 31; DB 1; Length 116;  
Best Local Similarity 85.7%; Pred. No. 39;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKRKA 7  
|||||:  
DB 49 KKKRKS 55

RESULT 10  
W04466  
ID W04466 standard; Protein; 14 AA.  
AC W04466;  
DT 30-JUL-1997 (first entry)  
DE Prenyl acceptor peptide moiety used in enzyme activity assay.  
KW Farnesyl transferase; inhibitor; cancer; tumour; neoplasia; prenyl;  
KW ras protein; K-ras B; malignant; detection; identification.  
OS Synthetic.  
PN W09634113-A2.  
PD 31-OCT-1996.  
PF 29-APR-1996; U05969.  
PR 27-APR-1995; US-429964.  
PA (TEXA ) UNIV TEXAS SYSTEM.  
PI Brown MS, Goldstein JL, James GL;

DR WPI: 96-497642/49.  
PT Assay for farnesyl transferase activity - by determining ability to  
PT transfer farnesyl moiety to K-Ras B protein, partic. useful for  
PT identifying inhibitors  
PS Example 4; Page 118; 257pp; English.  
CC W04466-W04468 are prenyl acceptor moieties which may be incorporated  
CC into the carboxyl terminals of proteins for the purpose of an enzyme  
CC activity assay, especially an assay for farnesyl transferase (FT)  
CC inhibitors. The FT enzyme can be used in a method for identifying FT  
CC inhibitors. The method involves screening candidate compounds for the  
CC ability to inhibit the transfer of a farnesyl moiety to a K-ras B  
CC protein. FT inhibitors act by blocking the attachment of prenyl groups  
CC to ras proteins in malignant cells of patients suffering from cancer  
CC or precancerous states, and as such are used to treat such conditions.  
SQ Sequence 14 AA;

Query Match 91.2%; Score 31; DB 1; Length 14;  
Best Local Similarity 85.7%; Pred. No. 6; 6;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKRKA 7  
|||||:  
DB 1 KKKRKS 7

RESULT 11  
W04468  
ID W04468 standard; Protein; 15 AA.  
AC W04468;  
DT 30-JUL-1997 (first entry)  
DE Prenyl acceptor peptide moiety used in enzyme activity assay.  
KW Farnesyl transferase; inhibitor; cancer; tumour; neoplasia; prenyl;  
KW ras protein; K-ras B; malignant; detection; identification.  
OS Synthetic.  
PN W09634113-A2.  
PD 31-OCT-1996.  
PF 29-APR-1996; U05969.  
PR 27-APR-1995; US-429964.  
PA (TEXA ) UNIV TEXAS SYSTEM.  
PI Brown MS, Goldstein JL, James GL;  
DR WPI: 96-497642/49.  
PT Assay for farnesyl transferase activity - by determining ability to  
PT transfer farnesyl moiety to K-Ras B protein, partic. useful for  
PT identifying inhibitors  
PS Example 4; Page 118; 257pp; English.  
CC W04466-W04468 are prenyl acceptor moieties which may be incorporated  
CC into the carboxyl terminals of proteins for the purpose of an enzyme  
CC activity assay, especially an assay for farnesyl transferase (FT)  
CC inhibitors. The FT enzyme can be used in a method for identifying FT  
CC inhibitors. The method involves screening candidate compounds for the  
CC ability to inhibit the transfer of a farnesyl moiety to a K-ras B  
CC protein. FT inhibitors act by blocking the attachment of prenyl groups  
CC to ras proteins in malignant cells of patients suffering from cancer  
CC or precancerous states, and as such are used to treat such conditions.  
SQ Sequence 15 AA;

Query Match 91.2%; Score 31; DB 1; Length 15;  
Best Local Similarity 85.7%; Pred. No. 7;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKRKA 7  
|||||:  
DB 1 KKKRKS 7

RESULT 12  
W04473  
ID W04473 standard; Protein; 188 AA.  
AC W04473;  
DT 05-AUG-1997 (first entry)  
DE Human K-ras B protein isoform.

KW Farnesyl transferase; inhibitor; cancer; tumour; neoplasia; prenyl;  
 KW ras protein; K-ras B; malignant; detection; identification.  
 OS Homo sapiens.  
 PN W0634113-A2.  
 PD 31-OCT-1996.  
 PF 25-APR-1996; U05969.  
 PR 27-APR-1995; US-429964.  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 PI Brown MS, Goldstein JL, James GL;  
 DR WPI: 96-497642/49.  
 PT Assay for farnesyl transferase activity - by determining ability to  
 transfer farnesyl moiety to K-Ras B protein, partic. useful for  
 PT identifying inhibitors  
 CC Example 5: Page 213-215; 257pp; English.  
 CS W04473 is one isoform of the human K-ras B protein. The human  
 CC K-ras B enzyme is alternatively spliced and the translated protein  
 CC may include one of two different exon IV regions. The K-ras B protein  
 CC was used in a method for identifying prenyl transferase inhibitors.  
 CC The method involved screening candidate compounds for the ability to  
 CC inhibit the transfer of a prenyl (farnesyl or geranylgeranyl) moiety  
 CC to a K-ras B protein. Inhibitors act by blocking the attachment  
 CC of prenyl groups to ras proteins in malignant cells of patients  
 CC suffering from cancer or precancerous states, and as such are used  
 CC to treat such conditions.  
 SQ Sequence 188 AA;

Query Match 91.2%; Score 31; DB 1; Length 188;  
 Best Local Similarity 85.7%; Pred. No. 59;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KKKKKKA 7  
 |||||:  
 Db 175 KKKKKKS 181

RESULT 13  
 ID W24559 standard; Protein: 1040 AA.  
 AC W24559;  
 DT 06-FEB-1998 (first entry)  
 DE Presentin-Interacting protein GT24.  
 KW Presentin-Interacting protein; human; Alzheimer's disease;  
 KW diagnosis; therapy; transgenic animal; animal model; GT24.  
 OS Homo sapiens.  
 FH Key  
 FT Domain location/Qualifiers  
 346..862  
 FT MISC-difference 12 /note="Presentin-Interacting domain"  
 FT MISC-difference 40 /note="encoded by GSR"  
 FT MISC-difference 71 /note="encoded by CCR"  
 FT MISC-difference 71 /note="encoded by AKC"  
 PN W09727296-A1.  
 PD 31-JUL-1997.  
 PF 27-JAN-1997; CA0051.  
 PR 02-JAN-1997; US-034590.  
 PR 26-JAN-1996; US-592541.  
 PR 05-JUL-1996; US-021673.  
 PR 12-JUL-1996; US-021700.  
 PR 08-NOV-1996; US-029895.  
 PA (HSCR-) HSC RES & DEV LP.  
 PI (UTOR ) UNIV TORONTO GOVERNING COUNCIL.  
 PI Fraser PE, Rommens JM, St George-Hyslop PH;  
 DR WPI: 97-393684/36.  
 DR N-PSDB: T79967.  
 PT Presentin-Interacting protein genes - used to develop products for  
 PT the diagnosis, therapy and study of Alzheimer's disease and related  
 PT disorders  
 CS Claim 1; Page 99-101; 133pp; English.  
 CC GT24, a human protein with 'armadillo' repeats, has been identified  
 CC as a presentin-Interacting protein (PIP). A yeast two-hybrid kit

CC was employed to screen a human brain cDNA library for clones which  
 CC interacted with presentin transmembrane 6-7 loop domain; mutations  
 CC in this loop domain are known to be causative of Alzheimer's  
 CC disease (AD). 9 PIP gene sequences (see T79966-74) including GT24  
 CC were identified. PIP nucleic acids, PIP proteins and peptides  
 CC (especially the presentin interacting domain), antibodies to PIPs,  
 CC cells transformed with PIP nucleic acids, and transgenic animals  
 CC altered with PIP nucleic acids can be used for the diagnosis,  
 CC therapy and study of AD and related disorders. They can be used to  
 CC identify compounds which can modulate the expression of a PIP gene  
 CC or which bind to a PIP or modulate its activity.  
 SQ Sequence 1040 AA;

Query Match 91.2%; Score 31; DB 1; Length 1040;  
 Best Local Similarity 85.7%; Pred. No. 2.5e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KKKKKKA 7  
 |||||:  
 Db 627 KKKKKKS 633

RESULT 14  
 ID W34654 standard; peptide: 17 AA.  
 AC W34654;  
 DT 01-APR-1998 (first entry)  
 DE C-terminal 17 amino acids of K-Ras-4B.  
 KW Raf protein; Ras; signal transduction; zinc finger domain;  
 KW Ras-Raf mediated signal transduction pathway; cell proliferation;  
 KW binding; inhibition; Raf activation inhibiting activity;  
 KW cell proliferation inhibition; immune cell receptor-mediated activation.  
 OS Unidentified.  
 PN W09734146-A1.  
 PD 18-SEP-1997.  
 PF 12-MAR-1997; U03881.  
 PR 11-MAR-1997; US-814836.  
 PR 12-MAR-1996; US-013274.  
 PA (GENO ) GEN HOSPITAL CORP.  
 PA (INDV ) UNIV INDIANA FOUND.  
 PI Avruch J, Luo Z, Marshall MS;  
 DR WPI: 97-470979/43.  
 PT Screening compounds which inhibit direct protein interaction -  
 PT useful to identify inhibitors of binding of Ras to Raf, Raf  
 PT activation and cell proliferation.  
 PS Disclosure; Page 12; 50pp; English.  
 CC The present sequence represents the C-terminal 17 amino acids of  
 CC K-Ras-4B. A wild type c-Raf-1 gene was altered to encode the present  
 CC sequence and a shortened N-terminal Myc epitope. These changes were made  
 CC to examine the other functions of the zinc finger domain of Raf. The zinc  
 CC finger domain binds to an epitope present only in prenylated Ras. The zinc  
 CC interaction participates in the transduction of an intracellular signal  
 CC via the Ras-Raf mediated signal transduction pathway which culminates in  
 CC cell proliferation. A novel method that inhibits the binding of Raf and  
 CC Ras is disclosed. The method is used to screen compounds for Raf  
 CC activation inhibiting activity, cell proliferation inhibition and  
 CC inhibition of direct binding of Ras to Raf. Blocking association of Ras  
 CC with Raf interferes with receptor-mediated activation of immune cells.  
 CC The method may also be useful in down-regulating the immune response in  
 CC patients with autoimmune diseases such as systemic lupus erythematosus,  
 CC type 1 diabetes, and rheumatoid arthritis.  
 SQ Sequence 17 AA;

Query Match 91.2%; Score 31; DB 1; Length 17;  
 Best Local Similarity 85.7%; Pred. No. 7.8;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KKKKKKA 7  
 |||||:  
 Db 4 KKKKKKS 10

## RESULT 15

W45898

ID W45898 standard; peptide; 214 AA.

DT 30-JUN-1998 (first entry)

DE SCR 1-3 of complement receptor type 1 (CR1)/switch fusion protein.

KW Membrane binding element; thrombotic disease; soluble protein;

KW complement-related disease; integral membrane protein; inflammation;

KW short consensus repeat; SCR 1-3; CR1; complement receptor type 1.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Cross\_links 214

FT /note= "Disulphide linked to Cys in peptide given

FT in W45898 or S-(CH2)2CONH(CH2)12CH3"

PN W09802454-A2.

PD 22-JAN-1998.

PF 08-JUL-1997; E03715.

PR 15-JUL-1996; GB-014871.

PA (ADPR-) ADPROTECH PLC.

PI Dodd I, Mossakowska DEI, Smith RAG;

DR WPI: 98-110524/10.

PT Derivatives of soluble poly:peptide(s) bonded to low affinity

PT membrane binding groups - useful for treating complement-related and

PT thrombotic diseases, providing improved localisation at cellular

PT membranes

PS Claim 22: Page 60: 75pp: English.

CC This sequence represents a specifically claimed protein having the

CC amino acid sequence of short consensus repeats (SCR 1-3) of

CC complement receptor type (ICRI) plus a switch fusion sequence. The

CC invention relates to a soluble derivative (A) of a soluble polypeptide

CC (I), which comprises at least 2 heterologous membrane-binding elements

CC (MBE) of low membrane affinity covalently associated with (I). MBE

CC interact, independently and with thermodynamic additivity, with

CC components of cellular or artificial membranes exposed to extracellular

CC fluids. (A) are used to treat disorders treatable with (I) itself,

CC specifically inflammation or any other complement-related disorder

CC (e.g. neurological disease, graft rejection, myocardial infarction,

CC sepsis, rheumatoid arthritis and many others; including application to

CC indwelling devices) and thrombotic disease, but also to treat allergy,

CC induce weight loss, to treat ischaemia or asthma and as immuno-

CC modulators for treating multiple sclerosis. (A) are administered orally,

CC topically, by injection or inhalation at 0.01-10 (preferably 0.1-10)

CC mg/kg/day.

SO Sequence 214 AA;

Query Match

Best Local Similarity 91.2%; Score 31; DB 1; Length 214;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKKA 7

DB 201 KKKKKKS 207

Search completed: July 18, 1999, 00:42:54  
 Job time: 8587 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 06:07:56 ; Search time 39.54 Seconds  
(without alignments)  
1.747 Million cell updates/sec

Title: US-09-142-043-6  
Perfect score: 34  
Sequence: 1 KKKKKKA 7

Scoring table: BLOSUM62

Searched: 106577 seqs, 9868381 residues

Database : Issued\_Patents\_AA:\*

- 1: /cgn2\_6/prodata/1/iaa/3A\_COMB.pep:\*
- 2: /cgn2\_6/prodata/1/iaa/3B\_COMB.pep:\*
- 3: /cgn2\_6/prodata/1/iaa/PCUT09\_COMB.pep:\*
- 4: /cgn2\_6/prodata/1/iaa/backfill.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	34	100.0	394	2	US-08-951-148-5	Sequence 5, Appl
2	34	100.0	394	2	US-08-951-148-10	Sequence 10, Appl
3	31	91.2	576	1	US-08-190-802A-56	Sequence 56, Appl
4	31	91.2	18	2	US-08-683-877-2	Sequence 2, Appl
5	30	88.2	11	1	US-07-694-983-15	Sequence 15, Appl
6	30	88.2	473	1	US-08-103-739B-2	Sequence 2, Appl
7	30	88.2	501	1	US-08-190-802A-28	Sequence 28, Appl
8	30	88.2	510	1	US-08-249-112-3	Sequence 3, Appl
9	30	88.2	487	1	US-08-249-112-4	Sequence 4, Appl
10	30	88.2	1213	1	US-08-188-582-20	Sequence 20, Appl
11	30	88.2	32	1	US-08-152-488-8	Sequence 8, Appl
12	30	88.2	29	1	US-08-054-363-17	Sequence 17, Appl
13	30	88.2	36	1	US-07-948-357-4	Sequence 4, Appl
14	30	88.2	36	1	US-07-948-357-12	Sequence 12, Appl
15	30	88.2	37	1	US-07-948-357-13	Sequence 13, Appl
16	30	88.2	37	1	US-08-231-730A-29	Sequence 29, Appl
17	30	88.2	32	1	US-08-303-025-8	Sequence 8, Appl
18	30	88.2	1235	1	US-08-118-101A-2	Sequence 2, Appl
19	30	88.2	1213	1	US-08-646-715-20	Sequence 20, Appl
20	30	88.2	29	1	US-08-088-658-22	Sequence 22, Appl
21	30	88.2	10	1	US-08-097-830E-1	Sequence 1, Appl
22	30	88.2	30	1	US-08-097-830E-2	Sequence 3, Appl
23	30	88.2	434	1	US-08-097-830E-3	Sequence 12, Appl
24	30	88.2	10	1	US-08-097-830E-12	Sequence 56, Appl
25	30	88.2	25	1	US-08-240-514-56	Sequence 2, Appl
26	30	88.2	1497	1	US-08-623-679-7	Sequence 7, Appl
27	30	88.2	1533	1	US-08-623-679-9	Sequence 9, Appl
28	30	88.2	398	1	US-08-507-431-2	Sequence 11, Appl
29	30	88.2	7	1	US-08-704-170-11	Sequence 29, Appl
30	30	88.2	37	1	US-08-427-001C-29	Sequence 29, Appl
31	30	88.2	32	1	US-08-677-304-8	Sequence 8, Appl
32	30	88.2	37	1	US-08-457-798-29	Sequence 29, Appl
33	30	88.2	691	2	US-08-405-648A-2	Sequence 2, Appl
34	30	88.2	37	2	US-08-457-171-29	Sequence 29, Appl
35	30	88.2	29	2	US-08-595-387-17	Sequence 17, Appl
36	30	88.2	148	2	US-08-468-347-22	Sequence 22, Appl
37	30	88.2	153	2	US-08-226-264-26	Sequence 26, Appl
38	30	88.2	182	2	US-08-226-264-28	Sequence 28, Appl
39	30	88.2	25	2	US-08-612-302A-56	Sequence 56, Appl

40	30	88.2	473	2	US-08-474-404-2	Sequence 2, Appl
41	30	88.2	148	2	US-08-467-389-22	Sequence 22, Appl
42	30	88.2	20	2	US-08-769-211-1	Sequence 1, Appl
43	30	88.2	20	2	US-08-769-211-4	Sequence 4, Appl
44	30	88.2	10	2	US-08-456-112B-1	Sequence 1, Appl
45	30	88.2	501	2	US-08-705-660-10	Sequence 10, Appl

## ALIGNMENTS

```

RESULT 1
US-08-951-148-5
; Sequence 5, Application US/08951148
; Patent No. 5871973
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti
APPLICANT: Shah, Purni
TITLE OF INVENTION: CELL DIVISION REGULATORS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/951.148
FILING DATE: Herewith
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0407 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
SEQUENCE CHARACTERISTICS:
LENGTH: 394 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: ZDANOT01
CLONE: 2458438
US-08-951-148-5

Query Match          100.0%; Score 34; DB 2; Length 394;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 KKKKKKA 7
DB      368 KKKKKKA 374

RESULT 2
US-08-951-148-10
; Sequence 10, Application US/08951148

```

Patent No. 5871973  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Bandman, Olga  
APPLICANT: Lal, Preeti  
APPLICANT: Shah, Purvi  
TITLE OF INVENTION: CELL DIVISION REGULATORS  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/951,148  
FILING DATE: Herewith  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0407 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 394 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 1167967  
US-08-951-148-10

Query Match 100.0%; Score 34; DB 2; Length 394;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKKA 7  
Db 368 KKKKKKA 374

RESULT 3  
US-08-190-802A-56  
Sequence 56, Application US/08190802A  
Patent No. 5519003  
GENERAL INFORMATION:  
APPLICANT: Mochly-Rosen, Daria  
APPLICANT: Ron, Dorit  
TITLE OF INVENTION: WD-40 - Derived Peptides and Uses  
NUMBER OF SEQUENCES: 265  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: P.O. Box 60850  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306-0850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/190,802A  
FILING DATE: 01-FEB-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Fadian, Gary R.  
REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 8600-0139  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 56:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 576 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Periodic Trp protein, Fig. 39  
US-08-190-802A-56

Query Match 91.2%; Score 31; DB 1; Length 576;  
Best Local Similarity 85.7%; Pred. No. 1.1e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKKA 7  
Db 270 KKKKKKS 276

RESULT 4  
US-08-683-877-2  
Sequence 2, Application US/08683877  
Patent No. 5776689  
GENERAL INFORMATION:  
APPLICANT: Karin, Michael  
APPLICANT: Ellledge, Stephen J.  
APPLICANT: Aronheim, Ami  
TITLE OF INVENTION: Protein Recruitment System  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell & Flores LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: United States  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/683,877  
FILING DATE: 19-JUL-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-UD 1971  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:



LENGTH: 18 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-683-877-2

Query Match 91.2%; Score 31; DB 2; Length 18;  
Best Local Similarity 85.7%; Pred. No. 6.9;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKK 7  
111111  
DB 5 KKKKKK 11

RESULT 5  
US-07-694-983-15

Sequence 15, Application US/07694983  
Patent No. 5432260

GENERAL INFORMATION:

APPLICANT: Stahl, Philip D.

TITLE OF INVENTION: HIGH AFFINITY MANNOS RECEPTOR

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Irell & Manella

STREET: 545 Middlefield Road, Suite 200

CITY: Menlo Park

STATE: California

COUNTRY: USA

ZIP: 94025

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/694,983

FILING DATE: 19910503

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Murashige, Kate H.

REGISTRATION NUMBER: 29,959

REFERENCE/DOCKET NUMBER: 9500-0039.00

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-327-7250

TELEFAX: 415-327-2951

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 11 amino acids

TYPE: AMINO ACID

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Peptide

LOCATION: 1

OTHER INFORMATION: /label= Ac-

FEATURE:

NAME/KEY: Peptide

LOCATION: 11

OTHER INFORMATION: /label= -NH2

US-07-694-983-15

Query Match 88.2%; Score 30; DB 1; Length 11;  
Best Local Similarity 100.0%; Pred. No. 6.7;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKK 6  
111111

DB 2 KKKKKK 7

RESULT 6

US-08-103-739B-2

Sequence 2, Application US/08103739B

Patent No. 5478369

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, Marc C.

APPLICANT: BEACH, Larry R.

APPLICANT: HOWARD, John A.

APPLICANT: HOFFMAN, Gary A.

TITLE OF INVENTION: DNA Sequences Mediating Male Fertility

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pioneer Hi-Bred International, Inc.

STREET: 700 Capital Square, 400 Locust Street

CITY: Des Moines

STATE: Iowa

COUNTRY: U.S.

ZIP: 50309

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/103,739B

FILING DATE: 02-AUG-1993

CLASSIFICATION: 800

PRIOR APPLICATION NUMBER: US 07/537,183

ATTORNEY/AGENT INFORMATION:

NAME: Sweeney, Patricia A.

REGISTRATION NUMBER: 32,733

REFERENCE/DOCKET NUMBER: 0125R2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (515) 248 4897

TELEFAX: (515) 248 4844

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 473 amino acids

TYPE: amino acid

TOPOLOGY: linear

US-08-103-739B-2

Query Match 88.2%; Score 30; DB 1; Length 473;

Best Local Similarity 100.0%; Pred. No. 1.4e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKK 6  
111111

DB 465 KKKKKK 470

RESULT 7

US-08-190-802A-28

Sequence 28, Application US/08190802A

Patent No. 5519003

GENERAL INFORMATION:

APPLICANT: Mochly-Rosen, Daria

APPLICANT: Ron, Dorit

TITLE OF INVENTION: WD-40 - Derived Peptides and Uses

NUMBER OF SEQUENCES: 265

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dehlinger & Associates

STREET: P.O. Box 60850

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94306-0850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/190,802A  
FILING DATE: 01-FEB-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Fabian, Gary R.  
REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 8600-0139  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 501 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Human 55 kda protein (PWP homolog),  
INDIVIDUAL ISOLATE: Fig. 11  
US-08-190-802A-28

Query Match 88.2%; Score 30; DB 1; Length 501;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KKKKK 6  
Db 240 KKKKK 245

RESULT 8  
US-08-249-112-3  
Sequence 3, Application US/08249112  
Patent No. 5527703  
GENERAL INFORMATION:  
APPLICANT: Cully, Doris F.  
APPLICANT: Arena, Joseph P.  
APPLICANT: Liu, Ken K.  
APPLICANT: Vassilatis, Demetrios  
TITLE OF INVENTION: DNA ENCODING GLUTAMATE GATED CHLORIDE  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wallen, John W.  
STREET: 126 E. Lincoln Ave., P.O. Box 2000  
CITY: Rahway  
STATE: New Jersey  
COUNTRY: U.S.A.  
ZIP: 07065  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/249,112  
FILING DATE: 25-MAY-1994  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Wallen, John W.  
REGISTRATION NUMBER: 35,403  
REFERENCE/DOCKET NUMBER: 19194  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (908) 594-3905  
TELEFAX: (908) 594-4720  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 510 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-249-112-3

Query Match 88.2%; Score 30; DB 1; Length 510;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KKKKK 6  
Db 505 KKKKK 510

RESULT 9  
US-08-249-112-4  
Sequence 4, Application US/08249112  
Patent No. 5527703  
GENERAL INFORMATION:  
APPLICANT: Cully, Doris F.  
APPLICANT: Arena, Joseph P.  
APPLICANT: Liu, Ken K.  
APPLICANT: Vassilatis, Demetrios  
TITLE OF INVENTION: DNA ENCODING GLUTAMATE GATED CHLORIDE  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wallen, John W.  
STREET: 126 E. Lincoln Ave., P.O. Box 2000  
CITY: Rahway  
STATE: New Jersey  
COUNTRY: U.S.A.  
ZIP: 07065  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/249,112  
FILING DATE: 25-MAY-1994  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Wallen, John W.  
REGISTRATION NUMBER: 35,403  
REFERENCE/DOCKET NUMBER: 19194  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908) 594-4720  
TELEFAX: (908) 594-3905  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 487 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-249-112-4

Query Match 88.2%; Score 30; DB 1; Length 487;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KKKKK 6  
Db 482 KKKKK 487

```
RESULT 10
US-08-188-582-20
; Sequence 20, Application US/08188582
; Patent No. 5534410
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Stegfried
; APPLICANT: Tanese, Neoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBRACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/188,582
; FILING DATE: 28-JAN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A.
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEEX: 910 277299
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1213 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-188-582-20

Query Match      88.2%; Score 30; DB 1; Length 1213;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KKKKK 6
      |||||
Db      1150 KKKKK 1155

RESULT 11
US-08-152-488-8
; Sequence 8, Application US/08152488
; Patent No. 5534619
; GENERAL INFORMATION:
; APPLICANT: Wakefield, Thomas W.
; APPLICANT: Andrews, Philip C.
; APPLICANT: Stanley, James C.
; TITLE OF INVENTION: NOVEL PEPTIDES FOR HEPARIN AND
; TITLE OF INVENTION: LOW MOLECULAR WEIGHT HEPARIN
; TITLE OF INVENTION: ANTICOAGULATION REVERSAL
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
```

```
ADDRESSEE: Benita J, Rohm, Esq.
STREET: 512 Springfield Avenue
CITY: Cranford
STATE: New Jersey
COUNTRY: United States of America
ZIP: 07016-1811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 6; ASCII (DOS)Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/152,488
; FILING DATE: 12-NOV-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/08069
; FILING DATE: 14-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Rohm, Benita J.
; REGISTRATION NUMBER: 28,664
; REFERENCE/DOCKET NUMBER: RM-7WG
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 908-276-3344
; TELEFAX: 908-276-5543
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 amino acids
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: N/A
; PUBLICATION INFORMATION:
; TITLE: N/A
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: PCT/US92/08069
; FILING DATE: 14-AUG-1993
; US-08-152-488-8

Query Match      88.2%; Score 30; DB 1; Length 32;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KKKKK 6
      |||||
Db      21 KKKKK 26

RESULT 12
US-08-054-363-17
; Sequence 17, Application US/08054363
; Patent No. 5539082
; GENERAL INFORMATION:
; APPLICANT: Nielsen, Peter E.
; APPLICANT: Buchardt, Ole
; APPLICANT: Egholm, Michael
; APPLICANT: Berg, Rolf H.
; TITLE OF INVENTION: No. 5539082el Peptide Nucleic Acids
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz, and No. 5539082ris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/054,363  
FILING DATE: 19930426  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Lucci, Joseph  
REGISTRATION NUMBER: 33,307  
REFERENCE/DOCKET NUMBER: ISIS1017  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-568-3100  
TELEFAX: 215-568-3439  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 29 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Cytosine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 2  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Cytosine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 3  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Cytosine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 4  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Cytosine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 5  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Cytosine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 6  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Adenine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 9 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 7  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Cytosine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 8  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /label= Modified-site

OTHER INFORMATION: /note= "Cytosine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 9  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Adenine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 9 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 10  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Cytosine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 11  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Thymine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 12  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Thymine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 13  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Cytosine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 14  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Cytosine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 15  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Cytosine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 16  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Cytosine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 17  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Thymine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 18  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Cytosine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."

FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 19  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Thymine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 20  
OTHER INFORMATION: /label= Modified-site  
OTHER INFORMATION: /note= "Cytosine heterocycle is attached to  
OTHER INFORMATION: N-acetyl(2-aminoethyl)glycine through the N-acetyl  
OTHER INFORMATION: group at position 1 of the heterocycle."  
US-08-054-363-17

Query Match 88.2%; Score 30; DB 1; Length 29;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KKKKK 6  
Db 21 KKKKK 26

RESULT 13  
US-07-948-357-4  
Sequence 4, Application US/07948357  
Patent No. 5547932  
GENERAL INFORMATION:  
APPLICANT: Curiel, David T.  
APPLICANT: Birnstiel, Max L.  
APPLICANT: Cotten, Matthew  
APPLICANT: Wagner, Ernst  
APPLICANT: Zatloukal, Kurt  
APPLICANT: Plank, Christian  
APPLICANT: Oberhauser, Berndt  
APPLICANT: Schmidt, Walter G.M.  
TITLE OF INVENTION: Composition for Introducing Nucleic Acid  
TITLE OF INVENTION: Complexes Into Higher Eucaryotic Cells  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1225 Connecticut Avenue  
City: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/948,357  
FILING DATE: 19920923  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Esmond, Robert W.  
REGISTRATION NUMBER: 32,893  
REFERENCE/DOCKET NUMBER: 0652,0940004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 466-0800  
TELEFAX: (202) 833-8716  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 36 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: both  
MOLECULE TYPE: peptide  
US-07-948-357-4

Query Match 88.2%; Score 30; DB 1; Length 36;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KKKKK 6  
Db 25 KKKKK 30

RESULT 14  
US-07-948-357-12  
Sequence 12, Application US/07948357  
Patent No. 5547932  
GENERAL INFORMATION:  
APPLICANT: Curiel, David T.  
APPLICANT: Birnstiel, Max L.  
APPLICANT: Cotten, Matthew  
APPLICANT: Wagner, Ernst  
APPLICANT: Zatloukal, Kurt  
APPLICANT: Plank, Christian  
APPLICANT: Oberhauser, Berndt  
APPLICANT: Schmidt, Walter G.M.  
TITLE OF INVENTION: Composition for Introducing Nucleic Acid  
TITLE OF INVENTION: Complexes Into Higher Eucaryotic Cells  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1225 Connecticut Avenue  
City: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/948,357  
FILING DATE: 19920923  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Esmond, Robert W.  
REGISTRATION NUMBER: 32,893  
REFERENCE/DOCKET NUMBER: 0652,0940004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 466-0800  
TELEFAX: (202) 833-8716  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 36 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: both  
MOLECULE TYPE: peptide  
US-07-948-357-12

Query Match 88.2%; Score 30; DB 1; Length 36;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KKKKK 6  
Db 27 KKKKK 32

RESULT 15  
US-07-948-357-13  
Sequence 13, Application US/07948357  
Patent No. 5547932

GENERAL INFORMATION:  
APPLICANT: Curitel, David T.  
APPLICANT: Birnstiel, Max L.  
APPLICANT: Colten, Matthew  
APPLICANT: Wagner, Ernst  
APPLICANT: Zalioukal, Kurt  
APPLICANT: Plank, Christian  
APPLICANT: Oberhauser, Berndt  
APPLICANT: Schmidt, Walter G.M.  
TITLE OF INVENTION: Composition for Introducing Nucleic Acid  
TITLE OF INVENTION: Complexes Into Higher Eucaryotic Cells  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1225 Connecticut Avenue  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/948,357  
FILING DATE: 19920923  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Esmond, Robert W.  
REGISTRATION NUMBER: 32,893  
REFERENCE/DOCKET NUMBER: 0652.0940004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 466-0800  
TELEFAX: (202) 833-8716  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 36 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: both  
MOLECULE TYPE: peptide  
US-07-948-357-13

Query Match 88.2%; Score 30; DB 1; Length 36;  
Best Local Similarity 100.0%; Pred.No. 17;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 KKKKK 6  
|||||  
Db 27 KKKKK 32

Search completed: July 18, 1999, 06:07:57  
Job time: 299 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 06:07:04 ; Search time 44.47 Seconds  
(without alignments)  
5.897 Million cell updates/sec

Title: US-09-142-043-6

Perfect score: 34

Sequence: 1 KKKKKKA 7

Scoring table: BLOSUM62

Searched: 116738 seqs, 37460341 residues

Database : PIR\_58:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	437	2	151238 translation elonga
2	34	100.0	1647	2	SNF2beta protein -
3	34	100.0	340	2	p38-2G4 protein -
4	34	100.0	1274	2	hypothetical prote
5	34	100.0	204	2	hypothetical prote
6	31	91.2	370	1	choleine-phosphate
7	31	91.2	1428	1	DNA topoisomerase
8	31	91.2	188	1	transforming prote
9	31	91.2	188	1	transforming prote
10	31	91.2	193	1	histone H5 - goose
11	31	91.2	43	1	small acid-soluble
12	31	91.2	188	2	transforming prote
13	31	91.2	188	2	transforming prote
14	31	91.2	188	2	transforming prote
15	31	91.2	194	2	histone H5 - musco
16	31	91.2	754	2	female sterile hom
17	31	91.2	285	2	translation initia
18	31	91.2	1085	2	hypothetical prote
19	31	91.2	295	2	hypothetical prote
20	31	91.2	587	2	hypothetical prote
21	31	91.2	142	2	hypothetical prote
22	31	91.2	590	2	hypothetical prote
23	31	91.2	2014	2	probable membrane
24	31	91.2	730	2	probable membrane
25	31	91.2	576	2	PWP1 protein - yea
26	31	91.2	146	2	signal recognition
27	31	91.2	147	2	high mobility grou
28	31	91.2	226	2	hypothetical prote
29	31	91.2	524	2	transcription fact
30	31	91.2	886	2	transcription term
31	31	91.2	1320	2	TCOF1 protein - mo
32	31	91.2	483	2	Nasopressin recept
33	30	88.2	533	1	monophenol monooxy
34	30	88.2	478	1	methionyl aminopep
35	30	88.2	154	1	ubiquitin / riboso
36	30	88.2	225	1	small nuclear ribo
37	30	88.2	199	1	potassium transpor
38	30	88.2	1235	1	heat shock protein
39	30	88.2	732	1	HHH086

40	30	88.2	733	1	HNMS86	heat shock protein
41	30	88.2	728	1	HCH90	heat shock protein
42	30	88.2	533	2	I49736	monophenol monooxy
43	30	88.2	1369	2	JC4860	protein-tyrosine k
44	30	88.2	480	2	A46702	methionyl aminopep
45	30	88.2	215	2	I52523	nucleoporin p62 ho

## ALIGNMENTS

## RESULT 1

151238 translation elongation factor EF-1 gamma - African clawed frog  
C:Species: Xenopus laevis (African clawed frog)  
C>Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 07-Aug-1998  
C/Accession: I51238  
R:Morales, J.; Bassez, T.; Cormier, P.; Mulner-Lorillon, O.; Belle, R.; Osborne, H.B.  
Dev. Genet. 14, 440-448, 1993  
A>Title: Expression of elongation factor 1 alpha (EF-1 alpha) and 1 beta gamma (EF-1  
A:Reference number: I51237; MUID:94155465  
A:Accession: I51238  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-437 <MOR>  
A:Cross-references: GB:S69726; NID:9545506; PID:9545507  
C:Superfamily: translation elongation factor eEF-1 gamma chain

Query Match 100.0%; Score 34; DB 2; Length 437;  
Best Local Similarity 100.0%; Pred. No. 33;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKKA 7  
DB 245 KKKKKKA 251

## RESULT 2

S45252 SNF2beta protein - human  
C:Species: Homo sapiens (man)  
C>Date: 10-Dec-1994 #sequence\_revision 17-Nov-1995 #text\_change 12-Sep-1997  
C/Accession: S45252  
R:Chiba, H.; Muramatsu, M.; Nomoto, A.; Kato, H.  
Nucleic Acids Res. 22, 1815-1820, 1994  
A>Title: Two human homologues of Saccharomyces cerevisiae SWI2/SNF2 and Drosophila br  
A:Reference number: S45251  
A:Accession: S45252  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1647 <CHT>  
C:Superfamily: unassigned bromodomain proteins; bromodomain homology  
F:1485-1540/Domain: bromodomain homology <BRO>

Query Match 100.0%; Score 34; DB 2; Length 1647;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKKA 7  
DB 583 KKKKKKA 589

## RESULT 3

S54181 p38-2G4 protein - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 08-Jul-1995 #sequence\_revision 21-Jul-1995 #text\_change 13-Sep-1998  
C/Accession: I48702; S54181  
R:Radomski, N.; Jost, E.  
Exp. Cell Res. 22, 434-445, 1995

A:Title: Molecular cloning of a murine cDNA encoding a novel protein, p38-2G4, which var  
A:Reference number: 148702  
A:Accession: 148702  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-340 <RES>  
A:Cross-references: EMBL:X84789; NID:g790469; PID:g790470  
C:Genetics:  
A:Gene: 2G4  
C:Superfamily: Schizosaccharomyces pombe 42K protein

Query Match 100.0%; Score 34; DB 2; Length 340;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 7  
|||  
Db 314 KKKKKKA 320

RESULT 4  
S28279  
hypothetical protein B0464.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 12-Mar-1993 #sequence\_revision 12-Mar-1993 #text\_change 30-Sep-1993  
C:Accession: S28279  
R:Kershaw, J.  
submitted to the EMBL Data Library, December 1992  
A:Reference number: S28278  
A:Accession: S28279  
A:Molecule type: DNA  
A:Residues: 1-1274 <KER>  
A:Cross-references: EMBL:Z19152  
C:Genetics:  
A:introns: 24/3; 162/3; 408/1; 465/3; 779/3; 1193/1; 1244/1

Query Match 100.0%; Score 34; DB 2; Length 1274;  
Best Local Similarity 100.0%; Pred. No. 81;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 7  
|||  
Db 1209 KKKKKKA 1215

RESULT 5  
B71433  
hypothetical protein - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana  
A:Variety: Columbia  
C:Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 03-Aug-1998  
C:Accession: B71433  
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk  
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; Giel  
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.  
Nature 391, 485-488, 1998  
A:Authors: Funk, B.; Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Pu  
H.; Duesterhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechma  
chuelier, C.; Chawatis, N.  
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal  
A:Reference number: A71400; MUID:98121113  
A:Accession: B71433  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-204 <BEV>  
A:Cross-references: GB:297341; NID:g2244991; PID:e327510; PID:g2245016  
C:Genetics:  
A:Map position: 4COP9-4G3845

Query Match 100.0%; Score 34; DB 2; Length 204;

Best Local Similarity 100.0%; Pred. No. 18;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 7  
|||  
Db 105 KKKKKKA 111

RESULT 6  
S68187  
choline-phosphate cytidyltransferase (EC 2.7.7.15) - Plasmodium falciparum  
N:Alternate names: CTP:choline-phosphate cytidyltransferase; phosphorylcholine tran  
C:Species: Plasmodium falciparum  
C:Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 26-Feb-1998  
C:Accession: S68187  
R:Iyo, H.J.; Sri Midada, J.; Mercereau-Puijalon, O.; Vidal, H.J.  
Eur. J. Biochem. 233, 62-72, 1995  
A:Title: Molecular cloning of CTP:phosphocholine cytidyltransferase from Plasmodium  
A:Reference number: S68187; MUID:96061933  
A:Accession: S68187  
A:Molecule type: DNA  
A:Residues: 1-370 <YEO>  
A:Cross-references: EMBL:X84041; NID:g1054826; PID:g1054827  
C:Genetics:  
A:Gene: CTP  
A:introns: #status absent  
C:Function:  
A:Description: catalyzes the formation of cytidinediphosphate choline and pyrophospha  
A:Pathway: phospholipid biosynthesis  
C:Superfamily: choline-phosphate cytidyltransferase  
C:Keywords: membrane-associated protein; nucleotidyltransferase; phospholipid biosynt  
F:274-294/Domain: amphipathic helix #status predicted <ANH>  
F:308-324/Domain: amphipathic helix #status predicted <ANH>

Query Match 91.2%; Score 31; DB 1; Length 370;  
Best Local Similarity 85.7%; Pred. No. 89;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 7  
|||  
Db 194 KKKKKKS 200

RESULT 7  
ISBYT2  
DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) - yeast (Saccharomyces cerevisiae)  
N:Alternate names: DNA gyrase; DNA topoisomerase II; protein YNL088W  
C:Species: Saccharomyces cerevisiae  
C:Date: 30-Sep-1992 #sequence\_revision 10-Nov-1995 #text\_change 12-Dec-1997  
C:Accession: S57534; A25630; S63027; S30866; S65093  
R:Solier-Mira, A.; Saiz, J.E.; Ballesta, J.P.G.; Remacha, M.  
submitted to the EMBL Data Library, June 1995  
A:Reference number: S57533  
A:Accession: S57534  
A:Molecule type: DNA  
A:Residues: 1-1428 <SOL>  
A:Cross-references: EMBL:X89016; NID:g887621; PID:g887623  
R:Glaeyer, G.; Lynn, R.; Goto, T.; Wang, J.C.  
J. Biol. Chem. 261, 12448-12454, 1986  
A:Title: The complete nucleotide sequence of the structural gene TOP2 of yeast DNA to  
A:Reference number: A25630; MUID:86304413  
A:Accession: A25630  
A:Molecule type: DNA  
A:Residues: 1-74, 'N', 75-546, 'L', 548-836, 'R', 838-1428 <GIA>  
A:Cross-references: GB:M13814; NID:g172997; PID:g172998  
R:Solier-Mira, A.; Saiz, J.E.; Ballesta, J.P.G.; Remacha, M.  
submitted to the Protein Sequence Database, April 1996  
A:Reference number: S63027  
A:Accession: S63027  
A:Molecule type: DNA  
A:Residues: 1-1428 <SOW>  
A:Cross-references: EMBL:Z71364; NID:g1301988; PID:e239691; PID:g1301989; MIPS:YNL088



A:Experimental source: strain S286C  
R:Jannatipour, M.; Liu, Y.X.; Nittiss, J.L.  
submitted to the EMBL Data Library, January 1993  
A:Description: The top2-5 mutant of yeast topoisomerase II encodes an enzyme resistant to  
A:Reference number: S30866  
A:Accession: S30866  
A:Molecule type: DNA  
A:Residues: 812-836, 'R', 838-882, 'P', 884, 'II', 887-977 <JAN>  
A:Cross-references: EMBL:L08968; NID:g172999; PID:g173000  
R:Soler-Mira, A.; Salz, J.E.; Ballesta, J.P.G.; Remacha, M.  
Yeast 12, 485-491, 1996  
A:Title: The sequence of a 17 933 bp segment of Saccharomyces cerevisiae chromosome XIV  
A:Reference number: S65092  
A:Accession: S65092  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-1428 <SOF>  
A:Cross-references: EMBL:X89016; NID:g887621; PID:g887623  
C:Comment: type II DNA topoisomerase catalyzes the ATP-dependent transient breakage, pas  
C:Genetics:  
A:Gene: SCD:TOP2; TOR3; TRF3; TOP2-5  
A:Cross-references: SCD:S0005032; MIPS:VNL088w  
A:Map position: 14L  
C:Superfamily: eukaryotic type II DNA topoisomerase  
C:Keywords: ATP; DNA binding; DNA replication; heterotetramer; isomerase; phosphoprotein

Query Match 91.2%; Score 31; DB 1; Length 1428;  
Best Local Similarity 85.7%; Pred. No. 2.7e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKKA 7  
|||||  
Db 333 KKKKKKS 339

RESULT 8  
TWO2K  
transforming protein K-ras-2, major splice form - human  
N:Alternate names: c-K-ras protein; c-Ki-ras protein; c-Kirsten-ras protein; p21 protein  
C:Species: Homo sapiens (man)  
C>Date: 19-Feb-1984 #sequence\_revision 19-Feb-1984 #text\_change 21-Aug-1998  
C:Accession: B93311; B93310; I59498; I51971; A01367  
R:McGarrath, J.P.; Capon, D.J.; Smith, D.H.; Chen, E.Y.; Seeburg, P.H.; Goeddel, D.V.; Ley  
Nature 304, 501-506, 1983  
A:Title: Structure and organization of the human Ki-ras proto-oncogene and a related pro  
A:Reference number: A93311; MUID:83271513  
A:Accession: B93311  
A:Molecule type: DNA  
A:Residues: 1-188 <MCG>  
A:Cross-references: GB:L00045; GB:L00046; GB:L00047; GB:L00049; NID:g190907; PID:g190910  
R:Shimizu, K.; Birbaum, D.; Ruley, M.A.; Fasano, O.; Suard, Y.; Edlund, L.; Taparowsky,  
Nature 304, 497-500, 1983  
A:Title: Structure of the Ki-ras gene of the human lung carcinoma cell line Calu-1.  
A:Reference number: A93310; MUID:83271512  
A:Accession: B93310  
A:Molecule type: DNA  
A:Residues: 1-11, 'C', 13-30, 'G', 32-188 <SHI>  
A:Cross-references: GB:L00049; GB:X00652; GB:X00653; NID:g190907  
R:Sanos, E.; Martin-Zanca, D.; Reddy, E.P.; Pierotti, M.A.; Della Porta, G.; Barbacid,  
Science 223, 661-664, 1984  
A:Title: Malignant activation of a K-ras oncogene in lung carcinoma but not in normal ti  
A:Reference number: I59498; MUID:8412180  
A:Accession: I59498  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-36 <SAN>  
A:Cross-references: GB:M34904; NID:g186765; PID:g553520  
R:Hirai, H.; Okabe, T.; Anraku, Y.; Fujisawa, M.; Urabe, A.; Takaku, F.  
Biochem. Biophys. Res. Commun. 127, 168-174, 1985  
A:Title: Activation of the c-K-ras oncogene in a human pancreas carcinoma.  
A:Reference number: I52205; MUID:85149400  
A:Accession: I52205

A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-60, 'H', 62-96 <HIR>  
A:Cross-references: GB:X03210; NID:g190899; PID:g553635  
A:Note: the sequence with a point mutation found in T3M-4 primary pancreatic exocrine  
R:Kahn, S.; Yamamoto, F.  
Anticancer Res. 7, 639-652, 1987  
A:Title: The c-K-ras gene and human cancer (review).  
A:Reference number: I51971; MUID:8802525  
A:Accession: I51971  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-11, 'C', 13-188 <KAR>  
A:Cross-references: GB:M54968; NID:g186763; PID:g186764  
C:Comment: This splice form, approximately 99% of the transcripts in colon carcinoma  
C:Genetics:  
A:Gene: GDB:KRAS2; K-ras  
A:Cross-references: GDB:120120; OMIM:190070  
A:Map position: 12p12.1-12p12.1  
A:introns: 37/3; 97/2; 150/3  
A:Note: the first intron occurs before the initiator codon  
C:Superfamily: ras transforming protein  
C:Keywords: alternative splicing; GTP binding; lipoprotein; membrane protein; methyl  
F:10-17/Region: nucleotide-binding motif A (P-loop)  
F:116-119/Region: GTP-binding NRXD motif  
F:145-147/Region: GTP-binding SAK/L motif  
F:16-17, 35, 116, 117, 119, 145/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #  
F:185/Modified site: methyl ester carboxyl end (Cys) (in mature form) #status predict  
F:185/Binding site: farnesyl (Cys) (covalent) #status predicted

Query Match 91.2%; Score 31; DB 1; Length 188;  
Best Local Similarity 85.7%; Pred. No. 51;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKKA 7  
|||||  
Db 175 KKKKKKS 181

RESULT 9  
TWS2K  
transforming protein K-ras, splice form 2 - mouse  
N:Alternate names: c-Ki-ras transforming protein  
C:Species: Mus musculus (house mouse)  
C>Date: 04-Dec-1986 #sequence\_revision 09-Apr-1998 #text\_change 21-Aug-1998  
C:Accession: B01365; I59500; A46134  
R:George, D.L.; Scott, A.F.; Trusko, S.; Glick, B.; Ford, E.; Dorney, D.J.  
EMBO J. 4, 1199-1203, 1985  
A:Title: Structure and expression of amplified cki-ras gene sequences in Y1 mouse adr  
A:Reference number: A01365; MUID:85230570  
A:Accession: B01365  
A:Molecule type: DNA  
A:Residues: 1-188 <GEO>  
A:Cross-references: GB:X02452; NID:g52798; PID:e224274; PID:g1213018  
A:Experimental source: adrenal tumor Y1  
A:Guerrero, I.; Villasante, A.; Corces, V.; Pellicer, A.  
Science 223, 1159-1162, 1984  
A:Title: Activation of a c-K-ras oncogene by somatic mutation in mouse lymphomas indu  
A:Reference number: I59500; MUID:84300296  
A:Accession: I59500  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-37 <GUE>  
A:Cross-references: GB:X01927; NID:g200673; PID:g200674  
R:You, M.; Wang, Y.; Stoner, G.; You, L.; Maronpot, R.; Reynolds, S.H.; Anderson, M.  
Proc. Natl. Acad. Sci. U.S.A. 89, 5804-5808, 1992  
A:Title: Parental bias of Ki-ras oncogenes detected in lung tumors from mouse hybrids  
A:Reference number: A46134; MUID:92335186  
A:Accession: A46134  
A:Status: translation not shown; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-37 <YOU>

A:Cross-references: GB:S39586; NID:g251108  
A:Experimental source: strains C3A and AC3 F1  
A:Note: the parental inbred strains, C3H/HeJ and A/J, are both Mus musculus  
C:Genetics:  
A:Gene: K-ras; Ki-ras  
A:Introns: 37/3; 97/2; 150/3  
A:Note: this gene is amplified in adrenal tumor Y1 cells  
C:Superfamily: ras transforming protein  
C:Keywords: alternative splicing; GTP binding; lipoprotein; membrane protein; methylated  
F:10-17/Region: nucleotide-binding motif A (P-loop)  
F:116-119/Region: GTP-binding NKXD motif  
F:145-147/Region: GTP-binding SAK/L motif  
F:16,17,35,116,117,119,145/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #sta  
F:185/Binding site: farnesyl (Cys) (covalent) #status predicted  
F:185/Modified site: methyl ester carboxyl end (Cys) (in mature form) #status predicted

Query Match 91.2%; Score 31; DB 1; Length 188;  
Best Local Similarity 85.7%; Pred. No. 51;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 7  
|||||  
Db 175 KKKKKKS 181

RESULT 10  
HSC55  
histone H5 - goose  
C:Species: Anser anser (domestic goose)  
C:Date: 31-Dec-1979 #sequence\_revision 31-Dec-1979 #text\_change 16-Feb-1997  
C:Accession: A02588  
R:Yaguchi, M.; Roy, C.; Seljig, V.L.  
Biochem. Biophys. Res. Commun. 90, 1400-1406, 1979  
A:Title: Complete amino acid sequence of goose erythrocyte H5 histone and the homology b  
A:Reference number: A02588; MUID:80087664  
A:Accession: A02588  
A:Molecule type: protein  
A:Residues: 1-193 <YAG>  
C:Superfamily: histone H1  
C:Keywords: chromosomal protein; DNA binding; erythrocyte  
F:1-21/Domain: amino-terminal <NH2>  
F:22-99/Domain: globular <GDB>  
F:100-193/Domain: carboxyl-terminal <END>

Query Match 91.2%; Score 31; DB 1; Length 193;  
Best Local Similarity 85.7%; Pred. No. 52;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 7  
|||||  
Db 109 KKKKKKA 115

RESULT 11  
A36939  
small acid-soluble spore protein SASP-G - Bacillus megaterium (strain QMB151) 111-kb p  
C:Species: Bacillus megaterium  
C:Date: 11-Nov-1994 #sequence\_revision 02-Dec-1994 #text\_change 05-Sep-1997  
C:Accession: A36939; B36939  
R:Carillo-Martinez, Y.; Setlow, P.  
J. Bacteriol. 175, 6337-6340, 1993  
A:Title: Cloning and nucleotide sequence of a plasmid-carried gene coding for a minor sm  
A:Reference number: A36939  
A:Accession: A36939  
A:Molecule type: DNA  
A:Residues: 1-43 <CAR>  
A:Cross-references: GB:L20464; NID:g304174; PID:g304175  
A:Note: authors translated the codon ACG for residue 33 as Ser  
A:Accession: B36939  
A:Molecule type: protein  
A:Residues: 2-41 <CAZ>

C:Comment: This minor small, acid-soluble spore protein (SASP) is unusual in containi  
C:Comment: No homolog to this protein could be detected in Bacillus subtilis or Bacil  
C:Genetics:  
A:Gene: ssfp  
C:Superfamily: small acid-soluble spore protein SASP-G  
C:Keywords: germination; sporulation; storage protein  
F:2-43/Product: small acid-soluble spore protein SASP-G #status experimental <MAT>

Query Match 91.2%; Score 31; DB 1; Length 43;  
Best Local Similarity 85.7%; Pred. No. 15;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 7  
|||||  
Db 19 KKKKKKS 25

RESULT 12  
I58402  
transforming protein K-ras - rat  
N:Alternate names: GTP-binding protein p21 (K-ras); Ki-ras  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 26-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 21-Aug-1998  
C:Accession: I58402; I57596; I52667; I59099  
R:Higbicham, K.G.; Rice, J.M.; Buzard, G.S.; Perantoni, A.O.  
Oncogene 9, 2455-2459, 1994  
A:Title: Activation of the K-ras gene by insertion mutations in chemically induced ra  
A:Reference number: I58402; MUID:94336198  
A:Accession: I58402  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-188 <RES>  
A:Cross-references: EMBL:009793; NID:g495533; PID:g495534  
R:Tahira, T.; Hayashi, K.; Ochiai, M.; Tsuchida, N.; Nagao, M.; Sugimura, T.  
Mol. Cell. Biol. 6, 1349-1351, 1986  
A:Title: Structure of the c-Ki-ras gene in a rat fibrosarcoma induced by 1, 8-dinitro  
A:Reference number: I57596; MUID:87064407  
A:Accession: I57596  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-11, 'C', 13-46, 'H', 48-96 <RE2>  
A:Cross-references: GB:M1260; NID:g206563; PID:g554503  
A:Experimental source: fibrosarcoma  
A:Note: this sequence represents a transforming mutant of K-ras with a defective nucl  
R:Irifani, A.; Katayama, N.; Tahira, T.; Hayashi, K.; Tsuchida, N.  
Bull. Tokyo Med. Dent. Univ. 33, 35-40, 1986  
A:Title: Nucleotide sequence of exon I of the rat c-K-ras gene.  
A:Reference number: I52667; MUID:86190265  
A:Accession: I52667  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-37 <RE3>  
A:Cross-references: GB:M54870; NID:g203481; PID:g203482  
R:McMahon, G.; Davis, E.; Wogan, G.N.  
Proc. Natl. Acad. Sci. U.S.A. 84, 4974-4978, 1987  
A:Title: Characterization of c-Ki-ras oncogene alleles by direct sequencing of enzym  
A:Reference number: I59099; MUID:87260911  
A:Accession: I59099  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-28 <REA>  
A:Cross-references: GB:M16970; NID:g205094; PID:g554465  
C:Genetics:  
A:Gene: K-ras  
A:Introns: 37/3  
A:Note: the list of introns is incomplete  
C:Superfamily: ras transforming protein  
C:Keywords: GTP binding; lipoprotein; membrane protein; methylated carboxyl end; onco  
F:10-17/Region: nucleotide-binding motif A (P-loop)  
F:116-119/Region: GTP-binding NKXD motif  
F:145-147/Region: GTP-binding SAK/L motif

F:16,17,35,116,117,119,145/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #sta  
F:165/Modified site: methyl ester carboxyl end (Cys) (in mature form) #status predicted  
F:185/Binding site: farnesyl (Cys) (covalent) #status predicted

## Query Match

91.2%; Score 31; DB 2; Length 188;

Best Local Similarity 85.7%; Pred. No. 51;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKKA 7

|||||:

Db 175 KKKKKKS 181

## RESULT 13

S31720

transforming protein (K-ras) - short-tailed opossum (Monodelphis domestica)

C:Species: Monodelphis domestica

C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 28-Aug-1998

C:Accession: S31720

R:Kusewitt, D.F.; Kelly, G.; Sabourin, C.L.K.; Ley, R.D.

Submitted to the EMBL Data Library, May 1992

A:Description: Cloning and sequencing of a cDNA derived from the K-ras proto-oncogene (K

A:Reference number: S31720

A:Accession: S31720

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-188 &lt;KUS&gt;

A:Cross-references: EMBL:Z12125

C:Superfamily: ras transforming protein

C:Keywords: GTP binding; P-loop

F:10-17/Region: nucleotide-binding motif A (P-loop)

F:116-119/Region: GTP-binding NKXD motif

F:145-147/Region: GTP-binding SAK/L motif

F:16,17,35,116,117,119,145/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #sta

## Query Match

91.2%; Score 31; DB 2; Length 188;

Best Local Similarity 85.7%; Pred. No. 51;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKKA 7

|||||:

Db 175 KKKKKKS 181

## RESULT 14

S34138

transforming protein (K-ras) - Kirsten murine sarcoma virus

C:Species: Kirsten murine sarcoma virus

C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 28-Aug-1998

C:Accession: S34138

R:Kim, S.; Park, S.; Park, J.

Submitted to the EMBL Data Library, June 1993

A:Description: Nucleotide sequence of a cDNA encoding the rat c-k-ras gene.

A:Reference number: S34138

A:Accession: S34138

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-188 &lt;KIM&gt;

A:Cross-references: EMBL:Z23152

C:Superfamily: ras transforming protein

C:Keywords: GTP binding; P-loop

F:10-17/Region: nucleotide-binding motif A (P-loop)

F:116-119/Region: GTP-binding NKXD motif

F:145-147/Region: GTP-binding SAK/L motif

F:16,17,35,116,117,119,145/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #sta

QY 1 KKKKKKA 7  
|||||:  
Db 175 KKKKKKS 181

## RESULT 15

S07260

histone H5 - muscovy duck

C:Species: Cairina moschata (muscovy duck)

C:Date: 12-Feb-1993 #sequence\_revision 12-Feb-1993 #text\_change 08-Sep-1997

C:Accession: S07260

R:Boencke, D.; Toenjes, R.

J. Mol. Biol. 178, 121-135, 1984

A:title: Conserved dyad symmetry structures at the 3' end of H5 histone genes. Analys

A:Reference number: S07260; MUID:85033692

A:Accession: S07260

A:Molecule type: DNA

A:Residues: 1-194 &lt;DOE&gt;

A:Cross-references: EMBL:X01065; NID:g62726; PID:g62727

C:Superfamily: histone H1

C:Keywords: chromosomal protein; DNA binding; nucleus

## Query Match

91.2%; Score 31; DB 2; Length 194;

Best Local Similarity 85.7%; Pred. No. 52;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKKA 7

|||||:

Db 110 KKKKKKA 116

Search completed: July 18, 1999, 06:07:05  
Job time: 251 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 11:37:52 ; Search time 31.53 Seconds  
(without alignments)  
5.959 Million cell updates/sec

Title: US-09-142-043-6

Perfect score: 34

Sequence: 1 KKKKKKA 7

Scoring table: BLOSUM62

Searched: 74019 segs, 26840295 residues

Database: SwissProt\_36.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	100.0	437	1 EF1H_XENLA	Q91375 xenopus lae
2	34	100.0	509	1 NA57_RAT	P40615 rattus norv
3	34	100.0	394	1 PASI_MOUSE	P50580 mus musculu
4	34	100.0	1647	1 SN24_HUMAN	P51532 homo sapien
5	34	100.0	1288	1 KY01_CAEEL	Q03560 caenorhabdi
6	31	91.2	421	1 AMP2_YEAST	P38174 saccharomyc
7	31	91.2	370	1 CTPT_PLAFL	P49587 plasmodium
8	31	91.2	193	1 H5_ANSAN	P02258 anser anser
9	31	91.2	193	1 H5_CAIMO	P06513 caltrina mos
10	31	91.2	310	1 IF2B_SCHPO	P56329 schizosacch
11	31	91.2	285	1 IF2B_YEAST	P09064 saccharomyc
12	31	91.2	576	1 PWP1_YEAST	P21504 saccharomyc
13	31	91.2	188	1 RASK_MONDO	Q07983 monodelphis
14	31	91.2	188	1 RASL_HUMAN	P01118 homo sapien
15	31	91.2	188	1 RASL_MOUSE	P08643 mus musculu
16	31	91.2	188	1 RASL_MOUSE	P46203 rattus norv
17	31	91.2	754	1 RAG3_HUMAN	P23440 homo sapien
18	31	91.2	146	1 SRI4_YEAST	P38985 saccharomyc
19	31	91.2	524	1 T2FA_XENLA	Q04870 xenopus lae
20	31	91.2	1428	1 TOR2_YEAST	P06786 saccharomyc
21	31	91.2	1085	1 YAF4_SCHPO	Q09863 schizosacch
22	31	91.2	2014	1 YUJ7_YEAST	P39326 saccharomyc
23	31	91.2	587	1 YKRO_YEAST	P34248 saccharomyc
24	31	91.2	142	1 YMG8_YEAST	Q03525 saccharomyc
25	31	91.2	590	1 YNM7_YEAST	P53563 saccharomyc
26	30	88.2	478	1 AMP2_HUMAN	P50579 homo sapien
27	30	88.2	478	1 AMP2_MOUSE	Q08663 mus musculu
28	30	88.2	478	1 AMP2_MOUSE	P38062 rattus norv
29	30	88.2	1332	1 B3A3_HUMAN	P48751 homo sapien
30	30	88.2	1227	1 B3A3_MOUSE	P23348 mus musculu
31	30	88.2	1227	1 B3A3_RAT	P23348 rattus norv
32	30	88.2	690	1 CNG1_BOVIN	Q00194 bos taurus
33	30	88.2	691	1 CNG1_CANFA	Q28279 canis famli
34	30	88.2	686	1 CNG1_HUMAN	P29873 homo sapien
35	30	88.2	684	1 CNG1_MOUSE	P29974 mus musculu
36	30	88.2	683	1 CNG1_RAT	Q62927 rattus norv
37	30	88.2	445	1 EF1G_YEAST	Q08649 saccharomyc
38	30	88.2	445	1 EF1G_YEAST	Q08649 saccharomyc
39	30	88.2	784	1 GCF_HUMAN	P16867 homo sapien
40	30	88.2	157	1 H2B3_VOLCA	P16867 volvox cart
41	30	88.2	155	1 H2B4_VOLCA	P16868 volvox cart
42	30	88.2	728	1 HS9A_CHICK	P11501 gallus gall
43	30	88.2	732	1 HS9A_CRIGR	P46633 cricetus

## ALIGNMENTS

```
RESULT 1
ID EF1H_XENLA STANDARD: PRT: 437 AA.
AC Q91375;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE ELONGATION FACTOR 1-GAMMA TYPE 2 (EF-1-GAMMA) (P47).
OC XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AMPHIBIA; ANURA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=OOCYTE;
RX MEDLINE: 94155465.
RA MORALES J., BASSETT T., CORMIER P., MULNER-LORILLON O., BELLE R.,
RA OSBORNE H.B.;
DEV. GENET. 14:440-448(1993).
CC -I- FUNCTION: PROBABLY PLAYS A ROLE IN ANCHORING THE COMPLEX TO OTHER
CC CELLULAR COMPONENTS.
CC -I- SUBUNIT: EF-1 IS COMPOSED OF FOUR SUBUNITS: ALPHA, BETA,
CC DELTA, AND GAMMA.
CC -I- SIMILARITY: TO OTHER SPECIES' GAMMA SUBUNIT AS WELL AS TO THE
CC N-TERMINAL OF SOME EUKARYOTIC AMINO ACID-tRNA SYNTHETASES.
DR EMBL: S69726; G545507; -.
DR PROSITE: PS50040: EF1G: 1.
KW ELONGATION FACTOR; PROTEIN BIOSYNTHESIS.
SQ SEQUENCE 437 AA; 50248 MW; 3D632B6E CRC32;
```

Query Match 100.0%; Score 34; DB 1; Length 437;  
Best local Similarity 100.0%; Pred. No. 19;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 KKKKKKA 7
Db 245 KKKKKKA 251

RESULT 2
ID NA57_RAT STANDARD: PRT: 509 AA.
AC P40615;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE NUCLEOLAR PROTEIN NAP57.
OC RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE: 95096165.
RA MEIER U., BLOBEL G.;
RL J. CELL BIOL. 127:1505-1514(1994).
RN [2]
RP REVISIONS TO C-TERMINAL.
RA MEIER U.;
RL SUBMITTED (DEC-1997) TO THE SWISS-PROT DATA BANK.
CC -I- SUBUNIT: ASSOCIATES WITH NOP140.
CC -I- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLAR.
CC -I- SIMILARITY: BELONGS TO THE NAP57/CBF5/TRUB FAMILY.
DR EMBL: Z34922; G550293; -.
KW NUCLEAR PROTEIN.
FT CONFLICT 459 509
ATPTTPRVKKKKKKKKADGGFEAEADGGDATTKKKKK
KARAAEELSG -> RRPAPRG (IN REF. 1).
```

SO SEQUENCE 509 AA; 56615 MW; 643EFF0F CRC32;

Query Match 100.0%; Score 34; DB 1; Length 509;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKA 7  
111111  
Db 495 KKKKKA 501

RESULT 3

PASL\_MOUSE STANDARD; PRT; 394 AA.

AC P50580;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE PROLIFERATION-ASSOCIATED PROTEIN 1 (PROTEIN P38-264).  
GN PLFAP.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NEF;  
RA NAKAGAWA Y., WATANABE S., AKIYAMA K., TSUTSUI K., INOUE H., SEKI S.;  
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE OF 55-394 FROM N.A.  
RC STRAIN=NEF;  
RX MEDLINE; 96032817.  
RA RADOMSKI N., JOSE E.;  
RL EXP. CELL RES. 220:434-445(1995).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN PROLIFERATING CELLS. OBSERVED  
CC BETWEEN G1 AND MID S PHASE, DECREASE TOWARD THE END OF S PHASE,  
CC AND DISAPPEAR AT THE S/G2 TRANSITION.  
CC -1- INDUCTION: BY MITOGENS.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M24C; ALSO KNOWN AS THE  
CC MAP FAMILY 2.

CC EMBL; U43918; G167967; -.  
DR EMBL; X84789; G790470; -.  
DR MGD; MGI:894684; PLFAP.  
DR PROSITE; PS01202; MAP\_2; 1.  
KW NUCLEAR PROTEIN; COBALT.  
FT METAL 109 109 COBALT (BY SIMILARITY).  
FT METAL 120 120 COBALT (BY SIMILARITY).  
FT METAL 188 188 COBALT (BY SIMILARITY).  
FT METAL 227 227 COBALT (BY SIMILARITY).  
FT METAL 321 321 COBALT (BY SIMILARITY).  
SO SEQUENCE 394 AA; 43699 MW; D813C4B8 CRC32;

Query Match 100.0%; Score 34; DB 1; Length 394;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKA 7  
111111  
Db 368 KKKKKA 374

RESULT 4

SN24\_HUMAN STANDARD; PRT; 1647 AA.

AC P51532;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE POSSIBLE GLOBAL TRANSCRIPTION ACTIVATOR SNF2L4 (SNF2-BETA) (BRG-1  
DE PROTEIN) (MITOTIC GROWTH AND TRANSCRIPTION ACTIVATOR) (BRAHMA

DE PROTEIN HOMOLOG 1).  
GN SNF2L4 OR BRG1 OR SNF2B.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 94050144.  
RA KHAVARI P.A., PETERSON C.L., TANKUN J.W., MENDEL D.B., CRABTREE G.R.;  
RL NATURE 366:170-174(1993).  
RN [2]  
RP REVISIONS.  
RA KHAVARI P.A., PETERSON C.L., TANKUN J.W., MENDEL D.B., CRABTREE G.R.;  
RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 94268902.  
RA CHIBA H., MORAMATSU M., NOMOTO A., KATO H.;  
RL NUCLEIC ACIDS RES. 22:1815-1820(1994).  
CC -1- FUNCTION: TRANSCRIPTIONAL COACTIVATOR COOPERATING WITH NUCLEAR  
CC HORMONE RECEPTORS TO POTENTIATE TRANSCRIPTIONAL ACTIVATION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.

CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.  
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.  
DR EMBL; U29175; G902046; -.  
DR PROSITE; PS00633; BROMODOMAIN\_1; 1.  
DR PROSITE; PS50014; BROMODOMAIN\_2; 1.  
KW TRANSCRIPTION REGULATION; NUCLEAR PROTEIN; ACTIVATOR; BROMODOMAIN;  
KW ATP-BINDING; HELICASE.  
FT DOMAIN 578 588 POLY-LYS.  
FT DOMAIN 663 672 POLY-GLU.  
FT NP\_BIND 779 786 ATP (POTENTIAL).  
FT SITE 881 884 DEGR. BOX.  
FT DOMAIN 1360 1364 POLY-GLU.  
FT DOMAIN 1477 1547 BROMODOMAIN.  
FT DOMAIN 1571 1584 POLY-GLU.  
SO SEQUENCE 1647 AA; 184585 MW; 6D22A23A CRC32;

Query Match 100.0%; Score 34; DB 1; Length 1647;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKA 7  
111111  
Db 583 KKKKKA 589

RESULT 5

YKDL\_CAEEL STANDARD; PRT; 1288 AA.

AC 003560;  
DT 01-FEB-1994 (REL. 28, CREATED)  
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 148.5 KD PROTEIN B0464.2 IN CHROMOSOME III.  
GN B0464.2.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE; 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COSEY T., COOPER J., COULSON A.,  
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,  
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,  
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N.,  
RA LATREILLE P., LIGHTNING J., LLOYD C., MORIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,  
RA SULLSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,

RA WATSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,  
RL WOHLDMAN P.;  
NA NATURE 368:32-38(1994).  
CC -1- SIMILARITY: TO YEAST CTR9.  
DR EMBL: Z19152; G1065679; -.  
DR PIR: S28279; S28279.  
DR WORMEP; B0464.2; CE00861.  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 1288 AA; 148490 MW; FC93CAC0 CRC32;

Query Match 100.0%; Score 34; DB 1; Length 1288;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKA 7  
|11111|  
Db 1223 KKKKKA 1229

## RESULT 6

AMP2\_YEAST  
ID AMP2\_YEAST STANDARD; PRT; 421 AA.  
AC P38174;  
DT 01-OCT-1994 (REL. 30, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE METHIONINE AMINOPEPTIDASE 2 (EC 3.4.11.18) (METAP 2) (PEPTIDASE M 2).  
GN MAP2 OR YBL091C OR YBL0701.  
OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=YPH500;  
RX MEDLINE: 96109265.  
RA LI X., CHANG Y.-H.;  
RL PROC. NATL. ACAD. SCI. U.S.A. 92:12357-12361(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C;  
RX MEDLINE: 96076635.  
RA OBERHAUSER B., GASSENHUBER J., PIRAVANDI E., DOMDEY H.;  
RL YEAST 11:1103-1112(1995).  
CC -1- FUNCTION: REMOVES THE AMINO-TERMINAL METHIONINE FROM NASCENT  
PROTEINS.  
CC -1- CATALYTIC ACTIVITY: L-METHIONYLPEPTIDE + H(2)O -> L-METHIONINE  
+ PEPTIDE.  
CC -1- COFACTOR: COBALT (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M24C; ALSO KNOWN AS THE  
MAP FAMILY 2.  
DR EMBL: U17437; G1045302; -.  
CC CC  
DR EMBL: X79489; G496684; ALT\_INIT.  
DR EMBL: Z35852; E304508; -.  
DR PIR: S45411; S45411.  
DR SGD: L0003044; MAP2.  
DR PROSITE: PS01202; MAP.2; 1.  
DR HYDROLASE; AMINOPEPTIDASE; COBALT.  
FT METAL 194  
FT METAL 205  
FT METAL 274  
FT METAL 307  
FT METAL 403  
FT METAL 403  
FT METAL 86  
FT METAL 86  
SQ SEQUENCE 421 AA; 47518 MW; 361C04E CRC32;

Query Match 91.2%; Score 31; DB 1; Length 421;  
Best Local Similarity 85.7%; Pred. No. 59;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKA 7  
|11111|  
Db 47 KKKKKS 53

## RESULT 7

CTPL\_PLAFK  
ID CTPL\_PLAFK STANDARD; PRT; 370 AA.  
AC P49387;  
DT 01-FEB-1996 (REL. 33, CREATED)  
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE CHOLINEPHOSPHATE CYTIDYLYLTRANSFERASE (EC 2.7.7.15) (PHOSPHORYLCHOLINE  
TRANSFERASE) (CT).  
GN CTP.  
OS PLASMODIUM FALCIPARUM (ISOLATE KI / THAILAND).  
OC EUKARYOTA; PROTOZOA; APICOMPLEXA; SPOROZOA; COCCIDIA; EUCCOCIDIIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 96061933.  
RA YEO H.-U., SRI WIDADA J., MERCEREAU-PUTALON O., VIAL H.J.;  
RL EUR. J. BIOCHEM. 233:62-72(1995).  
CC -1- FUNCTION: CONTROLS PHOSPHATIDYCHOLINE SYNTHESIS.  
CC -1- CATALYTIC ACTIVITY: CTP + CHOLINE PHOSPHATE -> PYROPHOSPHATE + CDP  
CHOLINE.  
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE ASEXUAL INTRACEROTROCYTIC  
STAGES.  
CC -1- SIMILARITY: BELONGS TO THE CHOLINEPHOSPHATE CYTIDYLYLTRANSFERASE  
FAMILY.  
DR EMBL: X64041; G1054827; -.  
CC CC  
KW TRANSFERASE; NUCLEOTIDYLYLTRANSFERASE; PHOSPHOLIPID BIOSYNTHESIS.  
FT DOMAIN 57  
FT DOMAIN 193  
FT DOMAIN 89  
FT DOMAIN 268  
SQ SEQUENCE 370 AA; 42630 MW; 8C89ABA5 CRC32;

Query Match 91.2%; Score 31; DB 1; Length 370;  
Best Local Similarity 85.7%; Pred. No. 52;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKA 7  
|11111|  
Db 194 KKKKKS 200

## RESULT 8

H5\_ANSAN  
ID H5\_ANSAN STANDARD; PRT; 193 AA.  
AC P02258;  
DT 21-JUL-1986 (REL. 01, CREATED)  
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1990 (REL. 16, LAST ANNOTATION UPDATE)  
DE HISTONE H5.  
OS ANSER ANSER ANSER (WESTERN GRAYLAG GOOSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;  
CC ANSERIFORMES.  
RN [1]  
RP SEQUENCE.  
RX MEDLINE: 80087664.  
RA YAGUCHI M., ROT C., SELIGY V.L.;  
RL BIOCHEM. BIOPHYS. RES. COMMON. 90:1400-1406(1979).  
CC -1- FUNCTION: HISTONE H5 PERFORMS THE SAME FUNCTION AS H1, BEING  
NECESSARY FOR THE CONDENSATION OF NUCLEOSOME CHAINS INTO HIGHER  
ORDER STRUCTURES. AND REPLACES HISTONE H1 IN CERTAIN CELLS.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- TISSUE SPECIFICITY: ERYTHROID CELLS.  
DR PIR: A02588; HSGS5.  
DR HSSP: P08287; 1H57.  
KW CHROMOSOMAL PROTEIN; NUCLEAR PROTEIN; DNA-BINDING; DNA CONDENSATION.  
FT DOMAIN 29  
FT DOMAIN 100  
SQ SEQUENCE 193 AA; 20900 MW; 24819DBF CRC32;

Query Match 91.2%; Score 31; DB 1; Length 193;  
Best Local Similarity 85.7%; Pred. No. 29;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKKKKA 7  
:|||||  
Db 109 RKKKKA 115

RESULT 9  
ID H5\_CAIMO STANDARD: PRT: 193 AA.  
AC P06513:  
DT 01-JAN-1988 (REL. 06, CREATED)  
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)  
DT 01-NOV-1990 (REL. 16, LAST ANNOTATION UPDATE)  
DE HISTONE H5.  
OS CAIRINA MOSCHATA (MUSCOVY DUCK).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;  
OC ANSERIFORMES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 85033692.  
RA DOENECKE D., TOENJES R.;  
J. MOL. BIOL. 178:121-135(1984).  
RL -1- FUNCTION: HISTONE H5 PERFORMS THE SAME FUNCTION AS H1, BEING  
NECESSARY FOR THE CONDENSATION OF NUCLEOSOME CHAINS INTO HIGHER  
ORDER STRUCTURES, AND REPLACES HISTONE H1 IN CERTAIN CELLS.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- TISSUE SPECIFICITY: ERYTHROID CELLS.  
DR EMBL: X01065; 662727; -.  
DR PIR: S07260; S07260.  
DR HSSP: P08287; 1HST.  
DR CHROMOSOMAL PROTEIN; NUCLEAR PROTEIN; DNA-BINDING; DNA CONDENSATION.  
FT INIT\_MET 0  
FT DOMAIN 23 101 GLOBULAR.  
SQ SEQUENCE 193 AA; 20874 MW; 45674E55 CRC32;

Query Match  
Best Local Similarity 91.2%; Score 31; DB 1; Length 193;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKKKKA 7  
:|||||  
Db 109 RKKKKA 115

RESULT 10  
ID IF2B\_SCHPO STANDARD: PRT: 310 AA.  
AC P56329;  
DT 15-JUL-1998 (REL. 36, CREATED)  
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE PROBABLE EUKARYOTIC TRANSLATION INITIATION FACTOR 2 BETA SUBUNIT (EIF-2-BETA) (FRAGMENT).  
GN TIE212 OR SPAC6B12.17C.  
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-972;  
RA GENTILES S., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WOOD V.;  
RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- FUNCTION: EIF-2 FUNCTIONS IN THE EARLY STEPS OF PROTEIN SYNTHESIS  
BY FORMING A TERNARY COMPLEX WITH GTP AND INITIATOR TRNA. THIS  
COMPLEX BINDS TO A 40S RIBOSOMAL SUBUNIT, FOLLOWED BY MRNA BINDING  
TO FORM A 43S PREINITIATION COMPLEX. JUNCTION OF THE 60S RIBOSOMAL  
SUBUNIT TO FORM THE 80S INITIATION COMPLEX IS PRECEDED BY  
HYDROLYSIS OF THE GTP BOUND TO EIF-2 AND RELEASE OF AN EIF-2-GDP  
CC BINARY COMPLEX. IN ORDER FOR EIF-2 TO RECYCLE AND CATALYZE ANOTHER  
CC ROUND OF INITIATION, THE GDP BOUND TO EIF-2 MUST EXCHANGE WITH GTP  
CC BY WAY OF A REACTION CATALYZED BY EIF-2B (BY SIMILARITY).  
CC -1- SUBUNIT: HETEROTRIMER COMPOSED OF AN ALPHA, A BETA, AND A GAMMA

CC CHAIN.  
CC -1- SIMILARITY: TO OTHER SPECIES EIF-2-BETA AND TO N-TERMINAL OF P67.  
DR EMBL: Z98531; E334107; -.  
KW INITIATION FACTOR; PROTEIN BIOSYNTHESIS; ZINC-FINGER.  
FT NON\_TER 1  
FT DOMAIN 9 15 POLY-LYS.  
FT DOMAIN 73 80 POLY-LYS.  
FT DOMAIN 105 112 POLY-LYS.  
FT ZN\_FING 262 286 POTENTIAL.  
SQ SEQUENCE 310 AA; 34743 MW; 3BEELFA9 CRC32;

Query Match  
Best Local Similarity 91.2%; Score 31; DB 1; Length 310;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKKKKA 7  
:|||||  
Db 107 RKKKKS 113

ID IF2B\_YEAST STANDARD: PRT: 285 AA.  
AC P09064;  
DT 01-NOV-1988 (REL. 09, CREATED)  
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE EUKARYOTIC TRANSLATION INITIATION FACTOR 2 BETA SUBUNIT (EIF-2-BETA).  
GN SUT3 OR TIE212 YPL237W.  
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 88311063.  
RA DONAHUE T.F., CIGAN A.M., PABICH E.K., VALAVICIUS B.C.;  
RL CELL 54:621-632(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C;  
RA POHL T.M.;  
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- FUNCTION: EIF-2 FUNCTIONS IN THE EARLY STEPS OF PROTEIN SYNTHESIS  
BY FORMING A TERNARY COMPLEX WITH GTP AND INITIATOR TRNA. THIS  
COMPLEX BINDS TO A 40S RIBOSOMAL SUBUNIT, FOLLOWED BY MRNA BINDING  
TO FORM A 43S PREINITIATION COMPLEX. JUNCTION OF THE 60S RIBOSOMAL  
SUBUNIT TO FORM THE 80S INITIATION COMPLEX IS PRECEDED BY  
HYDROLYSIS OF THE GTP BOUND TO EIF-2 AND RELEASE OF AN EIF-2-GDP  
CC BINARY COMPLEX. IN ORDER FOR EIF-2 TO RECYCLE AND CATALYZE ANOTHER  
CC ROUND OF INITIATION, THE GDP BOUND TO EIF-2 MUST EXCHANGE WITH GTP  
CC BY WAY OF A REACTION CATALYZED BY EIF-2B.  
CC -1- SUBUNIT: HETEROTRIMER COMPOSED OF AN ALPHA, A BETA, AND A GAMMA  
CHAIN.  
CC -1- SIMILARITY: TO OTHER SPECIES EIF-2-BETA AND TO N-TERMINAL OF P67.  
DR EMBL: M21813; G171450; -.  
DR EMBL: Z67751; G1061252; -.  
DR EMBL: Z73594; E246947; -.  
DR PIR: S29368; S29368.  
DR SGD: I0002179; SUT3.  
KW INITIATION FACTOR; PROTEIN BIOSYNTHESIS; ZINC-FINGER.  
FT DOMAIN 16 23 LYS-RICH (BASIC).  
FT DOMAIN 49 56 LYS-RICH (BASIC).  
FT DOMAIN 82 89 LYS-RICH (BASIC).  
FT ZN\_FING 236 262 C4-TYPE (POTENTIAL).  
SQ SEQUENCE 285 AA; 31574 MW; 620BFF71 CRC32;

Query Match  
Best Local Similarity 91.2%; Score 31; DB 1; Length 285;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKKKKA 7  
:|||||



DB 49 KKKKKK 55

RESULT 12

PMPL\_YEAST

ID PMPL\_YEAST STANDARD; PRT; 576 AA.

AC P21304;

DT 01-MAY-1991 (REL. 18, CREATED)

DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)

DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

DE PERIODIC TRYPTOPHAN PROTEIN 1.

GN PMPL OR YLR196W OR L8167.10.

OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST).

OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-S288C / AB972;

RX MEDLINE: 92279208.

RA DUBONIO R.J., GORDON J.I., BOGUSKI M.S.;

RL PROTEINS 13:41-56(1992).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-S288C / AB972;

RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DU Z., FAVELLO A., FULTON L., GATTUNG S., GRECO T., KIRSTEN J., KUCABA T., HALLSWORTH K., LANGSTON Y., LATREILLE P., MAROIS E., MENZES S., MILLER N., NHAN M., PAULEY A., PELUSO D., RIKKEN L., RILES L., TATICH S., TREVASKIS E., VAUDIN M., VIGNATI D., WILCOX L., WILSON R., WOHLDMAN P., WATERSTON R.;

RL SUBMITTED (SEP-1994) TO EMBL/GENBANK/DBJ DATA BANKS.

CC -1- FUNCTION: NOT KNOWN.

CC -1- SIMILARITY: CONTAINS 8 WD REPEATS (TRP-ASP DOMAINS).

CC -1- SIMILARITY: BELONGS TO THE PMPL FAMILY OF WD-REPEAT PROTEINS.

DR EMBL: M25738; G172309.

DR EMBL: U14913; G544507.

DR PIR: S29367; S29367.

DR PIR: S48549; S48549.

DR SGD: L0001539; PMPL.

DR PROSITE: PS00678; WD\_REPEATS. 3.

DR REPEAT: WD REPEAT.

DR KW REPEAT: WD REPEAT.

SQ SEQUENCE 576 AA: 63803 MW: 2B32E42A CRC32;

Query Match 91.2%; Score 31; DB 1; Length 576;  
Best Local Similarity 85.7%; Pred. No. 78;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKK 7

DB 270 KKKKKK 276

RESULT 13

RASK\_MONDO

ID RASK\_MONDO STANDARD; PRT; 188 AA.

AC 007983;

DT 01-NOV-1995 (REL. 32, CREATED)

DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)

DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)

DE TRANSFORMING PROTEIN P21/K-RAS.

GN KRAS.

OS MONODOLPHIS DOMESTICA (SHORT-TAILED GREY OPOSSUM).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-EYE;

RA MEDLINE: 94146407.

RX KUSEWIT D.F., KELLY G., SABOURIN C.L.K., LEY R.D.;

RL DNA SEQ. 4:37-42(1993).

CC -1- FUNCTION: RAS PROTEINS BIND GDP/GTP AND POSSESS INTRINSIC GTPASE ACTIVITY.

CC -1- DISEASE: MUTATION WHICH CHANGES AA 61 IS IMPLICATED IN ULTRAVIOLET RADIATION-INDUCED (UVR-INDUCED) EYE TUMOR.

DR EMBL: 212125; G556894.

KW PROTO-ONCOGENE: GTP-BINDING; PRENIATION; LIPOPROTEIN; DISEASE MUTATION.

FT NP\_BIND 10 17 GTP (BY SIMILARITY).

FT NP\_BIND 57 61 GTP (BY SIMILARITY).

FT NP\_BIND 116 119 GTP (BY SIMILARITY).

FT DOMAIN 32 40 EFFECTOR REGION (BY SIMILARITY).

FT DOMAIN 166 185 HYPERVARIABLE REGION.

FT LIPID 185 185 FARNESYL (BY SIMILARITY).

FT VARIANT 61 61 Q -> L (IN UVR-INDUCED CORNEAL TUMOR).

SQ SEQUENCE 188 AA: 21396 MW: 581ABCE CRC32;

Query Match 91.2%; Score 31; DB 1; Length 188;  
Best Local Similarity 85.7%; Pred. No. 29;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKK 7

DB 175 KKKKKK 181

RESULT 14

RASL\_HUMAN

ID RASL\_HUMAN STANDARD; PRT; 188 AA.

AC P01118;

DT 21-JUL-1986 (REL. 01, CREATED)

DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE TRANSFORMING PROTEIN P21/K-RAS 2B.

GN KRAS2.

OS HOMO SAPIENS (HUMAN)

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; PRIMATES.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 83271513.

RA MCGRATH J.P., CAPON D.J., SMITH D.H., CHEN E.Y., SEEBURG P.H., GOEDEL D.V., LEVINSON A.D.;

RL NATURE 304:501-506(1983).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE-LUNG CARCINOMA;

RX MEDLINE: 83271512.

RA SHIMIZU K., BIRNBAUM D., RULEY M.A., FASANO O., SUARD Y., EDLUND L., TAPAROWSKY E., GOLDFARB M., WIGLER M.;

RL NATURE 304:497-500(1983).

RN [3]

RP SEQUENCE FROM N.A.

RC TISSUE-LUNG, AND COLON CARCINOMA;

RX MEDLINE: 83271514.

RA CAPON D.J., SEEBURG P.H., MCGRATH J.P., HAYFLICK J.S., EDMAN U., LEVINSON A.D., GOEDEL D.V.;

RL NATURE 304:507-513(1983).

RN [4]

RP SEQUENCE FROM N.A.

RC TISSUE-COLON CARCINOMA;

RX MEDLINE: 85036297.

RA MCCOY M.S., BARGMANN C.I., WEINBERG R.A.;

RL MOL. CELL. BIOL. 4:1577-1582(1984).

RN [5]

RP SEQUENCE FROM N.A.

RX MEDLINE: 88022525.

RA KAHN S., YAMAMOTO F., ALMOGUERA C., WINTER E., FORRESTER K., JORDANO J., PERUCHO M.;

RL ANTICANCER RES. 7:639-652(1987).

RN [6]

RP SEQUENCE OF 1-37 FROM N.A.

RC TISSUE-LUNG CARCINOMA;

RX MEDLINE: 84119465.

RA NAKANO H., YAMAMOTO F., NEVILLE C., EVANS D., MIZUNO T., PERUCHO M.;

PROC. NATL. ACAD. SCI. U.S.A. 81:71-75(1984).  
[7]  
RN SEQUENCE OF 1-37 FROM N.A.  
RX MEDLINE; 87215620.  
RA DENG G., LU Y., CHEN S., MIAO J., LU G., LI H., CAI H., XU X.,  
RA ZHENG E., LIU P.,  
RL CANCER RES. 47:3195-3198(1987).  
[8]  
RN SEQUENCE OF 1-36 FROM N.A.  
RX TISSUE-LUNG CARCINOMA;  
MEDLINE; 84121280.  
RA SANTOS E., MARTIN-ZANCA D., REDDY P.E., PIEROTTI M.A., PORTA G.,  
RA BARBACID M.,  
RL SCIENCE 223:661-664(1984).  
[9]  
RN SEQUENCE OF 38-96 FROM N.A.  
RX TISSUE-LUNG CARCINOMA;  
MEDLINE; 85087906.  
RA YAMAMOTO F., PERUCHO M.,  
RL NUCLEIC ACIDS RES. 12:8873-8885(1984).  
[10]  
RN SEQUENCE OF 1-96 FROM N.A.  
RX TISSUE-PANCREATIC CARCINOMA;  
MEDLINE; 85149400.  
RA HIRAI H., OKABE T., ANRAKU Y., FUJISAWA M., URABE A., TAKAKU F.,  
RL BIOCHEM. BIOPHYS. RES. COMMUN. 127:168-174(1985).  
[11]  
RN REVIEW;  
RX MEDLINE; 87297453.  
RA BARBACID M.,  
RL ANNU. REV. BIOCHEM. 56:779-827(1987).  
CC -1- FUNCTION: RAS PROTEINS BIND GDP/GTP AND POSSESS INTRINSIC GTPASE  
ACTIVITY.  
CC -1- THE NORMAL HUMAN SEQUENCE IS SHOWN.  
CC -1- THE MAMMALIAN RAS GENE FAMILY CONSISTS OF THE HARVEY AND KIRSTEN  
RAS GENES (C-HRAS1 AND C-KRAS2), AN INACTIVE PSEUDOGENE OF  
EACH (C-HRAS2 AND C-KRAS1) AND THE N-RAS GENE.  
CC -1- ALTERNATIVE PRODUCTS: THE C-TERMINAL PART OF THE 2 KRAS2 PROTEINS  
ARE CODED BY TWO ALTERNATIVE EXONS (IVA AND IVB).  
DR EMBL; L00049: G190910; -;  
DR EMBL; L00045: G190910; JOINED.  
DR EMBL; L00046: G190910; JOINED.  
DR EMBL; L00047: G190910; JOINED.  
DR EMBL; K01519: -; NOT\_ANNOTATED\_CDS.  
DR EMBL; K01520: -; NOT\_ANNOTATED\_CDS.  
DR EMBL; X01669: E3874; -;  
DR EMBL; X02825: E3875; -;  
DR EMBL; M34968: G186764; -;  
DR EMBL; M35876: G180475; -;  
DR EMBL; M30539: G190938; -;  
DR EMBL; M34904: G553520; -;  
DR EMBL; K03210: G553635; -;  
DR EMBL; K03209: G553635; JOINED.  
DR PIR; A01367; TVH02K.  
DR HSSP; P01112; 1PLK.  
DR MIM; 190070; -;  
KM PHOTO-ONCOGENE: GTP-BINDING; PRENYLATION; LIPOPROTEIN;  
KW ALTERNATIVE SPLICING; DISASE MUTATION.  
FT NP\_BIND 10 17 GTP.  
FT NP\_BIND 57 61 GTP.  
FT NP\_BIND 116 119 GTP.  
FT DOMAIN 32 40 EFFECTOR REGION.  
FT DOMAIN 166 185 HYPERVARIABLE REGION.  
FT LIPID 185 185 FARNESYL.  
FT VARIANT 12 12 G-> C (IN LUNG CARCINOMA).  
FT VARIANT 12 12 G-> V (IN COLON CARCINOMA).  
FT VARIANT 61 61 Q-> H (IN LUNG CARCINOMA PR310 AND  
PANCREAS T3M-4).  
FT MUTAGEN 164 164 R->A: LOSS OF GTP-BINDING ACTIVITY.  
FT SEQUENCE 188 AA: 21424 MM; ADAC1121 CRC32;

Best Local Similarity 85.7%; Pred. No. 29;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KKKKKA 7  
|||||:  
Db 175 KKKKKK 181  
RESULT 15  
RSL\_MOUSE  
ID RSL\_MOUSE STANDARD; PRT; 188 AA.  
AC P08643; P04200;  
DT 13-AUG-1987 (REL. 05, CREATED)  
DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE TRANSFORMING PROTEIN P21/K-RAS 2B.  
GN KRAS2.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.  
RN [1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE; 85230570.  
RA GEORGE D.L., SCOTT A.F., TRUSKO S., GLICK B., FORD E., DORNEY D.J.,  
RL EMBO J. 4:1199-1203(1985).  
[2]  
RN SEQUENCE OF 1-37 FROM N.A.  
RX MEDLINE; 84300296.  
RA GUERRERO I., VILLASANTE A., CORCES V., PELLICER A.,  
RL SCIENCE 225:1159-1162(1984).  
[3]  
RN SEQUENCE OF 109-173 FROM N.A.  
RX MEDLINE; 84169569.  
RA GEORGE D.L., SCOTT A.F., DE MARTINVILLE B., FRANCKE U.,  
RL NUCLEIC ACIDS RES. 12:2731-2743(1984).  
[4]  
RN REVIEW.  
RX MEDLINE; 87297453.  
RA BARBACID M.,  
RL ANNU. REV. BIOCHEM. 56:779-827(1987).  
CC -1- FUNCTION: RAS PROTEINS BIND GDP/GTP AND POSSESS INTRINSIC GTPASE  
ACTIVITY.  
CC -1- THE NORMAL MOUSE SEQUENCE IS SHOWN.  
CC -1- THIS GENE IS AMPLIFIED IN THE MOUSE ADRENAL TUMOR Y1 CELLS.  
CC -1- THE MAMMALIAN RAS GENE FAMILY CONSISTS OF THE HARVEY AND KIRSTEN  
RAS GENES (C-H-RAS1 AND C-K-RAS2), AN INACTIVE PSEUDOGENE OF  
EACH (C-H-RAS2 AND C-K-RAS1) AND THE N-RAS GENE.  
CC -1- ALTERNATIVE PRODUCTS: THE C-TERMINAL PART OF THE 2 KRAS2 PROTEINS  
ARE CODED BY TWO ALTERNATIVE EXONS (IVA AND IVB).  
DR EMBL; X02452: E224274; -;  
DR EMBL; X02453: E224274; JOINED.  
DR EMBL; X02454: E224274; JOINED.  
DR EMBL; X02456: E224274; JOINED.  
DR EMBL; K01927: G200674; -;  
DR EMBL; X00485; G929681; -;  
DR HSSP; P01112; 1PLK.  
DR MGD; MGI:96680; KRAS2.  
KM PHOTO-ONCOGENE: GTP-BINDING; PRENYLATION; LIPOPROTEIN;  
KW ALTERNATIVE SPLICING.  
FT NP\_BIND 10 17 GTP.  
FT NP\_BIND 57 61 GTP.  
FT NP\_BIND 116 119 GTP.  
FT DOMAIN 32 40 EFFECTOR REGION.  
FT DOMAIN 166 185 HYPERVARIABLE REGION.  
FT LIPID 185 185 FARNESYL.  
FT MUTAGEN 164 164 R->A: LOSS OF GTP-BINDING ACTIVITY.  
FT SEQUENCE 188 AA: 21482 MM; E46E355B CRC32;  
Query Match 91.2%; Score 31; DB 1; Length 188;  
Best Local Similarity 85.7%; Pred. No. 29;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKA 7  
| | | | |  
Db 175 KKKKKS 181

Search completed: July 18, 1999, 11:37:52  
job time: 1751 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 06:02:50 ; Search time 63.5 seconds

(without alignments)  
6.082 Million cell updates/sec

Title: US-09-142-043-6

Perfect score: 34

Sequence: 1 KKKKKKA 7

Scoring table: BLOSUM62

Searched: 180763 segs, 55169189 residues

Database :

SPREMBL\_8:\*  
1: sp\_fungi:\*  
2: sp\_human:\*  
3: sp\_invertebrate:\*  
4: sp\_mammal:\*  
5: sp\_mhc:\*  
6: sp\_organelle:\*  
7: sp\_phage:\*  
8: sp\_plant:\*  
9: sp\_bacteria:\*  
10: sp\_protent:\*  
11: sp\_virus:\*  
12: sp\_vertebrate:\*  
13: sp\_unclassified:\*  
14: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	34	100.0	394	2	043846	043846 homo sapien
2	34	100.0	1465	3	017909	017909 caenorhabd
3	34	100.0	2437	3	077393	077393 plasmodium
4	34	100.0	204	8	023504	023504 arabidopsis
5	34	100.0	149	10	035327	035327 mus musculu
6	34	100.0	392	12	073806	073806 fugu rubrip
7	34	100.0	1630	12	090753	090753 gallus gall
8	31	91.2	295	1	012515	012515 saccharomyc
9	31	91.2	860	1	036021	036021 schizosacch
10	31	91.2	426	1	060085	060085 schizosacch
11	31	91.2	730	1	099248	099248 saccharomyc
12	31	91.2	72	1	P89493	P89493 saccharomyc
13	31	91.2	886	2	015361	015361 homo sapien
14	31	91.2	876	2	043840	043840 homo sapien
15	31	91.2	754	2	015310	015310 homo sapien
16	31	91.2	1225	2	000379	000379 homo sapien
17	31	91.2	1030	2	076039	076039 homo sapien
18	31	91.2	683	3	009977	009977 caenorhabd
19	31	91.2	147	3	025870	025870 plasmodium
20	31	91.2	226	3	021198	021198 caenorhabd
21	31	91.2	597	3	002137	002137 caenorhabd
22	31	91.2	861	3	021571	021571 caenorhabd
23	31	91.2	368	3	045198	045198 caenorhabd
24	31	91.2	850	3	024211	024211 drosophila
25	31	91.2	927	3	044070	044070 urostylea gr
26	31	91.2	557	3	022243	022243 caenorhabd
27	31	91.2	617	3	044570	044570 caenorhabd
28	31	91.2	162	8	065509	065509 arabidopsis
29	31	91.2	925	8	065662	065662 arabidopsis

## ALIGNMENTS

30	31	91.2	267	8	Q39092	Q39092 arabidopsis
31	31	91.2	139	8	Q43371	Q43371 allium cepa
32	31	91.2	685	8	Q41382	Q41382 spinacia ol
33	31	91.2	431	8	O80587	O80587 arabidopsis
34	31	91.2	1336	8	O81074	O81074 arabidopsis
35	31	91.2	43	9	Q45320	Q45320 bacillus me
36	31	91.2	1302	10	O08857	O08857 mus musculu
37	31	91.2	1320	10	O08784	O08784 mus musculu
38	31	91.2	264	10	O35116	O35116 rattus norv
39	31	91.2	483	10	O63035	O63035 rattus norv
40	31	91.2	798	10	O54795	O54795 mus musculu
41	31	91.2	1247	10	O35927	O35927 mus musculu
42	31	91.2	798	10	O88411	O88411 mus musculu
43	31	91.2	188	11	O82954	O82954 kirsten mur
44	31	91.2	733	12	O73897	O73897 gallus gall
45	31	91.2	188	12	O93280	O93280 pleuronecte

RESULT 1  
043846 ID 043846 PRELIMINARY: PRT: 394 AA.  
AC 043846;  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE CELL CYCLE PROTEIN P38-264 HOMOLOG.  
GN HG4-1.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 98005911.  
RA LAMARTINE J., SERI M., HEITZMANN F., CREAVEN M., RADOMSKI N.,  
RA JOST E., LENOIR G.M., ROMEO G., SYLLA B.S.;  
RT "Molecular cloning and mapping of a human cDNA (PA2G4) that encodes a  
RT protein highly homologous to the mouse cell cycle protein p38-264";  
RL CYTOGENET. CELL GENET. 78:31-35(1997).  
DR EMBL: 059435; G2697005; -;  
DR PROSITE: PS01202; MAP\_2; 1.  
SQ SEQUENCE 394 AA; 43813 MW; A796DED5 CRC32;

Query Match 100.0%; Score 34; DB 2; Length 394;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKKA 7  
DB 368 KKKKKKA 374  
RESULT 2  
017909 ID 017909 PRELIMINARY: PRT: 1465 AA.  
AC 017909;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE H06001.2  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;  
OC RHABDITIDA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA BARLOW K.;  
RT SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 94150718.

RA WILSON R., AINSOUCHE R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,  
 RA BORTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,  
 RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,  
 RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSEHAW J.,  
 RA KIRSTEIN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,  
 RA MCNURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCIC C.,  
 RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,  
 RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,  
 RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
 RA WILKINSON-SPROAT J., WOHLDMAN P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans.";  
 RL NATURE 368:32-38(1994).  
 DR EMBL: Z92970; E349623; -;  
 DR PROSITE: PS00598; CHROMO.1; 2.  
 DR PFAM: PF00176; SNF2.N; 1.  
 DR PFAM: PF00271; helicase\_C; 1.  
 DR PFAM: PF00385; chromo; 2.  
 SQ SEQUENCE 1465 AA; 169391 MW; F959D52F CRC32;

Query Match 100.0%; Score 34; DB 3; Length 1465;  
 Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 7  
 |||||  
 DB 1015 KKKKKKA 1021

RESULT 3  
 077393 PRELIMINARY; PRT; 2437 AA.  
 AC 077393;  
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE MAL3P6.2 PROTEIN.  
 GN MAL3P6.2.  
 OS PLASMODIUM FALCIPARUM.  
 OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RA MURPHY L., LAMSON D., BARRELL B.;  
 RL SUBMITTED (OCT-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: Z98551; E1331931; -;  
 SQ SEQUENCE 2437 AA; 295348 MW; E6280382 CRC32;

Query Match 100.0%; Score 34; DB 3; Length 2437;  
 Best Local Similarity 100.0%; Pred. No. 7.4e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 7  
 |||||  
 DB 1391 KKKKKKA 1397

RESULT 4  
 023504 PRELIMINARY; PRT; 204 AA.  
 AC 023504;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE HYPOTHEICAL 22.8 KD PROTEIN.  
 OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).  
 OC EUKARYOTA; VIRIDIPHYTES; CHAROPHYTA/EMBRYOPHYTA GROUP; EMBRYOPHYTA;  
 OC TRACHEOPHYTA; EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA;  
 OC EUDICOTYLEDONS; ROSIDAE; CAPRARALES; BRASSICACEAE; ARABIDOPSIS.  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;  
 RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 98121113.  
 RA BEVAN M., BANCROFT I., BENT E., LOVE K., GOODMAN H., DEAN C.,  
 RA BEVGKAMP R., DIRKSE W., VAN STAVEREN M., STIEREMA W., DROST L.,  
 RA RIDLEY P., HUDSON S.A., PATEL K., MURPHY G., PIFANELLI P., WEDLER H.,  
 RA WEDLER E., WAMBUTT R., WEITZENEGER T., POHL T.M., TERRY N.,  
 RA GELTER J., VILARROEL R., DE CLEERCK R., VAN MONTAGU M., LECHARNY A.,  
 RA ADBORG S., GT I., KREIS M., LAO N., KAVANAGH T., HEMPEL S., KOTTER P.,  
 RA ENTIAN K.D., RIEGER M., SCHAEFER W., FUNK B., MUELLER-AUER S.,  
 RA SILVER M., JAMES R., MONTFORT A., PONS A., PIGDOMECH P., DOUKA A.,  
 RA HILBERTATOU E., MILIONI D., HATZIOPOULOS P., PIRAVANDI E., OBERMAIER B.,  
 RA HILBERT H., DUESTERHOFF A., MOORES T., JONES J.D.G., ENVA T.,  
 RA PLUME K., BENES V., RECHMAN S., ANSGORGE W., COOKE R., BERGER C.,  
 RA DELSENY M., VOET M., VOLCKAERT G., MEMES H.W., KLOSTERMAN S.,  
 RA SCHUELLER C., CHALWATZIS N.;  
 RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of  
 Arabidopsis thaliana.";  
 RL NATURE 391:485-488(1998).  
 DR EMBL: Z97341; E327510; -;  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2; 1.  
 DR PFAM: PF00096; ZF-C2H2; 1.  
 KW HYPOTHETICAL PROTEIN; ZINC-FINGER; METAL-BINDING; DNA-BINDING.  
 SQ SEQUENCE 204 AA; 22837 MW; 6CA3FDDA CRC32;

Query Match 100.0%; Score 34; DB 8; Length 204;  
 Best Local Similarity 100.0%; Pred. No. 80;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 7  
 |||||  
 DB 105 KKKKKKA 111

RESULT 5  
 035327 PRELIMINARY; PRT; 149 AA.  
 AC 035327;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE PROLINE-RICH PROTEIN 7 (FRAGMENT).  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCIUROGNATHI; MURIDAE; MORINAE; MUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA ERMEKOVA K., CHANG A., ZAMBRANO N., DE CANDIA P., RUSSO T., SUDOL M.;  
 RL ADV. EXP. MED. BIOL. 0:0-0(1997).  
 DR EMBL: AF020311; G2624968; -;  
 FT NON\_TER 1  
 FT NON\_TER 149  
 SQ SEQUENCE 149 AA; 16993 MW; 6F24B5EF CRC32;

Query Match 100.0%; Score 34; DB 10; Length 149;  
 Best Local Similarity 100.0%; Pred. No. 61;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 7  
 |||||  
 DB 139 KKKKKKA 145

RESULT 6  
 073806 PRELIMINARY; PRT; 392 AA.  
 AC 073806;  
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)

DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE PASI.  
OS FUGU RUBRIPES (JAPANESE PUFFERFISH) (TAKIFUGU RUBRIPES).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;  
OC TELEOSTEI; EUTELEOSTEI; ACANTHOPTERYGII; PERCOMORPHA;  
OC TETRAODONTIFORMES; TETRAODONTIDEI; TETRAODONTIDAE; FUGU.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA GELLNER K., BRENNER S.;  
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: AF056116; G3170545; -  
DE PROSITE: PS01202; MAP-2; 1.  
SQ SEQUENCE 392 AA; 43626 MW; 209BD93C CRC32;

Query Match 100.0%; Score 34; DB 12; Length 392;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKK 7  
DB 368 KKKKKK 374

RESULT 7  
ID 090753 PRELIMINARY; PRT; 1630 AA.  
AC 090753;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DE BRG1 PROTEIN.  
GN BRG1.  
OS GALLUS GALLUS (CHICKEN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;  
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=HAEMOPOIETIC;  
RX MEDLINE: 97169142.  
RA GOOWIN G.H.;  
RT "Isolation of cDNAs encoding chicken homologues of the yeast SNF2 and  
RT Drosophila Brahma proteins";  
RL GENE 184:27-32(1997).  
DR EMBL: X91637; G996018; -  
DR PROSITE: PS00633; BROMODOMAIN\_1; 1.  
DR PFAM: PF00116; SNF2\_N; 1.  
DR PFAM: PF00271; helicase\_C; 1.  
DR PFAM: PF00439; bromodomain; 1.  
SQ SEQUENCE 1630 AA; 183420 MW; 2F089105 CRC32;

Query Match 100.0%; Score 34; DB 12; Length 1630;  
Best Local Similarity 100.0%; Pred. No. 5.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKK 7  
DB 581 KKKKKK 587

RESULT 8  
ID 012515 PRELIMINARY; PRT; 295 AA.  
AC 012515;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DE ORE YDL173W.  
OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCAROMYCETALES;  
OC SACCAROMYCETACEAE; SACCAROMYCES.  
RN [1]

RP SEQUENCE FROM N.A.  
RA POHL T.M.;  
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA MIPS;  
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ALPHAS28C;  
RA POHL T.M.;  
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: Z74221; E253081; -  
DR EMBL: Z67750; G1061265; -  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 295 AA; 31886 MW; F9388DD3 CRC32;

Query Match 91.2%; Score 31; DB 1; Length 295;  
Best Local Similarity 85.7%; Pred. No. 3.1e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKK 7  
DB 278 KKKKKK 284

RESULT 9  
ID 036021 PRELIMINARY; PRT; 860 AA.  
AC 036021;  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DE HYPOTHETICAL 98.0 KD PROTEIN C4F10.09C IN CHROMOSOME I.  
GN SPAC4F10.09C.  
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;  
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;  
OC SCHIZOSACCHAROMYCES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RA CONNOR R., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WOOD V.;  
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -!- SIMILARITY: TO YEAST YDR060W.  
DR EMBL: Z98980; E340027; -  
KW HYPOTHETICAL PROTEIN.  
FT DOMAIN 67 70 POLY-LEU.  
FT DOMAIN 835 840 POLY-LYS.  
SQ SEQUENCE 860 AA; 97992 MW; 658F17BE CRC32;

Query Match 91.2%; Score 31; DB 1; Length 860;  
Best Local Similarity 85.7%; Pred. No. 8e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKK 7  
DB 835 KKKKKK 841

RESULT 10  
ID 060085 PRELIMINARY; PRT; 426 AA.  
AC 060085;  
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DE METHIONINE AMINOPEPTIDASE.  
GN SPBC14C8.03.  
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;

OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;  
 OC SCHIZOSACCHAROMYCETES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972H-;  
 RA LYNE M., RAJANDREAM M.A., BARRELL B.G., VOICKAERT G.;  
 RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: AL022305; E1285395; -  
 DR PROSITE: PS01202; MAP\_2; 1.  
 KM AMINOPEPTIDASE.  
 SO SEQUENCE 426 AA; 47271 MW; A80E0013 CRC32;

Query Match 91.2%; Score 31; DB 1; Length 426;  
 Best Local Similarity 85.7%; Pred. No. 4.3e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 7  
 DB 49 KKKKKKS 55

RESULT 11  
 ID 099248 PRELIMINARY; PRT; 730 AA.  
 AC 099248;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE CHROMOSOME XY READING FRAME ORF YOR019M.  
 OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCCHAROMYCETALES;  
 OC SACCCHAROMYCETACEAE; SACCCHAROMYCETES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA DE HAAN M., GRIVELL L.A., MAARSE A.C.;  
 RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MIPIS;  
 RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-FY1679;  
 RA DE HAAN M., MAARSE A.C., GRIVELL L.A.;  
 RL SUBMITTED (MAY-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-FY1679;  
 RA DUMONT M.E., SCHLICHTER J.B., CARDILLO T.S., HAYES M.K., BETHLENDY G.,  
 RA SHERMAN F.;  
 RT "CYC2 encodes a factor involved in mitochondrial import of yeast  
 cytochrome c.";  
 RL MOL. CELL. BIOL. 13:6442-6451(1993).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-FY1679;  
 RX MEDLINE: 94169519.  
 RA LEE Y.S., SHIMIZU J., YODA K., YAMASAKI M.;  
 RT "Molecular cloning of a gene, DHS1, which complements a  
 drug-hypersensitive mutation of the yeast *Saccharomyces cerevisiae*.";  
 RL BIOSCI. BIOTECHNOL. BIOCHEM. 58:391-395(1994).  
 DR EMBL: Z74927; E252323; -  
 DR EMBL: X87331; G829129; -  
 SO SEQUENCE 730 AA; 83565 MW; 09D72CD9 CRC32;

Query Match 91.2%; Score 31; DB 1; Length 730;  
 Best Local Similarity 85.7%; Pred. No. 6.9e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 7

DB 717 KKKKKKS 723

RESULT 12  
 ID P89493 PRELIMINARY; PRT; 72 AA.  
 AC P89493;  
 DT 01-MAY-1997 (TREMBLREL. 03, CREATED)  
 DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE ORF YBL091C (FRAGMENT).  
 GN MAP2.  
 OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCCHAROMYCETALES;  
 OC SACCCHAROMYCETACEAE; SACCCHAROMYCETES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA DOMDEY H., GASSENHUBER H., OBERMAIER B., PIRAVANDI E.;  
 RL SUBMITTED (SEP-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MIPIS;  
 RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 95112788.  
 RA FELDAMANN H., ATILE M., ALJINOVIC G., ANDRE B., BACLET M.C., BARTHE C.,  
 RA BAIR A., BECAM A.M., BITEAU N., BOLES E., BRANDT T., BRENDL M.,  
 RA BROECKNER M., BUSSEBERG F., CHRISTIANSEN C., CONTRERAS R., CROUZET M.,  
 RA CIEPLOUCH C., DEMOLIS N., DELAVEAU T., DOIGNON F., DOMEY H.,  
 RA DIESTERHUS S., DUBOIS E., DUJON B., EL BAKOURY M., ENTIAN K.D.,  
 RA FEUERMAN M., FIERIS W., FOBO G.M., FRITZ C., GASSENHUBER H.,  
 RA GLANSDORFF N., GOFREAU A., GRIVELL L.A., DE HAAN M., HEIN C.,  
 RA HERBERT C.J., HOLLENBERG C.P., HOLMSTROM K., JACO C., JACQUET M.,  
 RA JAUNIAUX J.C., JONNIAUX J.L., KALLESOE T., KIESAU P., KIRCHRAH L.,  
 RA KETTER P., KOROL S., LIBEL S., LOGCHE M., LOHAN A.J.E., LOUIS E.J.,  
 RA LI Z.Y., MAAT M.J., MALLEY L., MANNHAUPT G., MESSENGUY F., MIOGA T.,  
 RA MOLEMAN S.F., MUELLER S., NASR F., OBERMAIER B., PERA J., PIERARD A.,  
 RA PIRAVANDI E., POHL F.M., POHL T.M., PORTER S., PROFT M., PURNELLE B.,  
 RA RAMEZANI RAD M., RIEGER M., ROSE M., SCHAFF-GERETENSCHAEGER I.,  
 RA SCHERENS B., SCHWARZLOSE C., SKALA J., SLOINSKI P.P., SMITS P.H.M.,  
 RA SOUCIET J.L., STEENSA H.Y., STUCKA R., URRESTARAZU A.,  
 RA VAN DER AART O.J., VAN DYCK L., VASAROTTI A., VETTER I.,  
 RA VIERENDEELS F., VISSERS S., WAGNER G., DE WENIGHOSE P., WOLFE K.H.,  
 RA ZAGUSKI M., ZIMMERMANN F.K., MEMES H.M., KLEINE K.;  
 RT "Complete DNA sequence of yeast chromosome II.";  
 RL EMBL: J35851; E304571; -  
 DR EMBL: Z35851; E304571; -  
 FT NON\_TER  
 SO SEQUENCE 72 AA; 8174 MW; E3699D12 CRC32;

Query Match 91.2%; Score 31; DB 1; Length 72;  
 Best Local Similarity 85.7%; Pred. No. 87;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 7  
 DB 47 KKKKKKS 53

RESULT 13  
 ID Q15361 PRELIMINARY; PRT; 886 AA.  
 AC Q15361;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE TRANSSCRIPTION FACTOR.  
 GN TTF-1.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;



OC CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE; 95320168.  
 RA EVERS R., GRUMWT I.;  
 RT "Molecular coevolution of mammalian ribosomal gene terminator  
 sequences and the transcription termination factor TTF-I";  
 RL PROC. NATL. ACAD. SCI. U.S.A. 92:5827-5831(1995).  
 DR EMBL; X83973; G639693; -.  
 KW TRANSCRIPTION TERMINATION.  
 SQ SEQUENCE 886 AA; 101218 MW; 93D6AEF5 CRC32;

Query Match 91.2%; Score 31; DB 2; Length 886;  
 Best Local Similarity 85.7%; Pred. No. 8.2e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 7  
 Db 220 KKKKKKS 226

RESULT 14

ID 043840 PRELIMINARY; PRT; 876. AA.  
 AC 043840;  
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE NEURAL PLAKOPHILIN RELATED ARM-REPEAT PROTEIN (FRAGMENT).  
 GN NPRAV.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
 OC CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE; 98002299.  
 RA PAFENHOLZ R., FRANK W.W.;  
 RT "Identification and localization of a neurally expressed member of the  
 plakoglobin/armadillo multigene family";  
 RL DIFFERENTIATION 61:293-304(1997).  
 DR EMBL; U52351; G2822195; -.  
 FT NON\_TER 1  
 SQ SEQUENCE 876 AA; 96442 MW; 02D752C6 CRC32;

Query Match 91.2%; Score 31; DB 2; Length 876;  
 Best Local Similarity 85.7%; Pred. No. 8.2e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 7  
 Db 463 KKKKKKS 469

RESULT 15

ID 015310 PRELIMINARY; PRT; 754 AA.  
 AC 015310; 000699; 000700;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE KINASE.  
 GN RING3.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
 OC CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA THORPE K.L., ABDULLA S., KAUFMAN J., TROMSDALE J., BECK S.;  
 RL IMMUNOGENETICS 0:0-0(0).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RA THORPE K.;  
 RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; X96670; E243292; -.  
 DR EMBL; 296104; E321404; -.  
 DR EMBL; 284497; E321426; -.  
 DR PROSITE; PS00633; BROMODOMAIN\_1; 2.  
 DR PFAM; PF00439; bromodomain; 2.  
 SQ SEQUENCE 754 AA; 83184 MW; 99ADC934 CRC32;

Query Match 91.2%; Score 31; DB 2; Length 754;  
 Best Local Similarity 85.7%; Pred. No. 7.1e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKKKKKA 7  
 Db 507 KKKKKKA 513

Search completed: July 18, 1999, 06:02:51  
 Job time: 3106 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 00:42:54 ; Search time 58.51 Seconds  
(without alignments)  
2.765 Million cell updates/sec

Title: US-09-142-043-7  
Perfect score: 50  
Sequence: 1 CGGRKVC 8

Scoring table: BLOSUM62

Searched: 162890 segs, 20225328 residues

Database: A\_Geneseq\_34:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	100.0	8	1 W31837	Peptide 3 from the
2	36	72.0	9	1 W25558	Synthetic ferritin
3	35	70.0	248	1 R13875	Murine Cytotoxic C
4	35	70.0	10	1 W10886	Mab anti-HBsAg bin
5	32	64.0	2317	1 P92219	CR1 protein New nu
6	32	64.0	18	1 R06052	Immunoreactive pep
7	32	64.0	2039	1 R1810	Human complement t
8	32	64.0	1537	1 R11982	Partial human comp
9	32	64.0	543	1 R28543	CR1-4 (35E, 37Y) a
10	32	64.0	543	1 R28544	CR1-4 (35E) analog
11	32	64.0	543	1 R28545	CR1-4 (37Y) analog
12	32	64.0	543	1 R28546	CR1-4 (44T, 47D, 4
13	32	64.0	543	1 R28547	CR1-4 (52S, 53S, 5
14	32	64.0	543	1 R28548	CR1-4 (57V, 59K) a
15	32	64.0	543	1 R28549	CR1-4 (64K, 65T) a
16	32	64.0	543	1 R28550	CR1-4 (64K) analog
17	32	64.0	543	1 R28551	CR1-4 (65T) analog
18	32	64.0	543	1 R28552	CR1-4 (78T, 79D) a
19	32	64.0	543	1 R28553	CR1-4 (85R, 87N) a
20	32	64.0	543	1 R28554	CR1-4 (92T, 94H) a
21	32	64.0	543	1 R28555	CR1-4 (92T) analog
22	32	64.0	543	1 R28556	CR1-4 (94H) analog
23	32	64.0	543	1 R28557	CR1-4 (99H, 103E)
24	32	64.0	543	1 R28558	CR1-4 (109N, 110A,
25	32	64.0	543	1 R28559	CR1-4 (114-117STK
26	32	64.0	543	1 R28560	CR1-4 (114S) analog
27	32	64.0	543	1 R28561	CR1-4 (115T) analog
28	32	64.0	543	1 R28562	CR1-4 (116K) analog
29	32	64.0	543	1 R28563	CR1-4 (117P) analog
30	32	64.0	543	1 R28564	CR1-4 (116K, 117P)
31	32	64.0	543	1 R28565	CR1-4 (121O) analog
32	32	64.0	481	1 R29091	CR1-4 (amino acids
33	32	64.0	481	1 R29092	CR1-4 (amino acids
34	32	64.0	543	1 R28566	CR1-4 (318R, 319N)
35	32	64.0	543	1 R28567	CR1-4 (318-321 RNP
36	32	64.0	543	1 R28568	CR1-4 (347T, 349Y)
37	32	64.0	543	1 R28569	CR1-4 (369-376 STR
38	32	64.0	543	1 R28570	CR1-4 (266-274 KTK
39	32	64.0	543	1 R28571	CR1-4 (364-367 NNA
40	32	64.0	2039	1 R36743	CR1. Nucleic acid
41	32	64.0	197	1 R47152	Sequence of solubl
42	32	64.0	76	1 R47153	Sequence of solubl
43	32	64.0	254	1 R47154	Sequence of solubl

## ALIGNMENTS

RESULT 1	44	32	64.0	254	1 R47155	Sequence of solubl
W31837	45	32	64.0	1930	1 W45899	Human complement r
ID W31837, standard: peptide; 8 AA.						
AC W31837;						
DE Peptide 3 from the short consensus repeat 3 of complement receptor 1.						
DT Peptide 3 from the short consensus repeat 3; long homologous repeat A; LHR-A;						
KW Short consensus repeat 3; long homologous repeat A; LHR-A;						
KW complement receptor 1; CR1; complement inhibition;						
KW anti-haemolytic activity; inflammation; thrombotic condition;						
KW inappropriate complement activation; ARDS; Alzheimer's disease.						
OS Synthetic.						
OS Homo sapiens.						
PN W09731944-A1.						
PD 04-FEB-1997.						
PF 26-FEB-1997; E00994.						
PR 02-MAR-1996; GB-004518.						
PA (SMIK ) SMITHKLINE BEECHAM PLC.						
PI Edge CM, Mossakowska DEL, Smith RAG;						
DR WPI: 97-446630/41.						
PT Peptide derived from short consensus repeat 3 of human complement						
PT receptor 1 between amino acids Cys154-Gly186, useful to treat						
PT disorder or disease associated with inflammation or inappropriate						
PT complement activation						
PS Claim 13; Page 25; 33PP; English.						
CC Synthetic soluble peptides W31835-38 are derived from the short						
CC consensus repeat 3 of the long homologous repeat A (LHR-A) of human						
CC complement receptor 1 (CR1). The present peptide is located between						
CC residues 159 and 164. These peptides have functional complement						
CC inhibitory, including anti-haemolytic, activity. Enhanced activity						
CC may be achieved by linking the peptides to a core structure						
CC (e.g. the MAP peptide) to produce a multimeric or chimeric polypeptide.						
CC This polypeptide can be used to treat a disorder or disease associated						
CC with inflammation or inappropriate complement activation. It can also be						
CC used to treat a thrombotic condition, adult respiratory distress syndrome						
CC (ARDS), wounds, Alzheimer's disease or a CNS inflammatory disorder, or						
CC delay hyperacute allograft or xenograft rejection.						
SO Sequence 8 AA;						
Query Match	100.0%;	Score 50;	DB 1;	Length 8;		
Best local Similarity	100.0%;	Pred. No. 1.3e+05;				
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;						
OY 1 CGGRKVC 8						
DB 1 CGGRKVC 8						
RESULT 2						
ID W25558						
AC W25558 standard: peptide; 9 AA.						
DT 10-NOV-1997 (first entry)						
DE Synthetic ferritin peptide #31.						
KW Functional surrogate; analyte; affinity receptor; immunoreactive group;						
KW mimic; homogenous immunoassay; detection; diagnostic analyte; Chlamydia;						
KW cardiac marker; tumour marker; allergen; hormone; fertility; myoglobin;						
KW pregnancy; infectious disease; ferritin; myosin light chain; tropoin;						
KW follicle stimulating hormone; human; growth hormone; immunoglobulin E;						
KW prolactin; parathyroid hormone; placental lactogen; hepatitis antigen;						
KW antibody; chorionic gonadotropin; luteinising hormone; cytomegalovirus;						
KW Streptococcus; rubella; toxoplasma; DK-MB; prostate-specific antigen;						
KW carcinoembryonic antigen; alpha-fetoprotein; herpes virus; CA125.						
OS Synthetic.						
KW Key						
FT disulfide_bond 2. .9 Location/Qualifiers						

PN W09641172-A1.  
PD 19-DEC-1996.  
PF 07-JUN-1996; U10498.  
PR 07-JUN-1995; US-476375.  
PA (CYTO-) CYTOGEN CORP.  
PI Carter JM, Lee-Own FV;  
DR WPI: 97-077284/07.  
PT Labelled functional surrogate of an analyte - useful as competitor  
PT molecule in affinity assays, esp. for detecting large macromolecules  
PS such as ferritin  
PS Claim 51; Page 27; 156pp; English.  
CC This sequence represents a synthetic ferritin peptide which was used as  
CC a functional surrogate in the conjugate of the invention. The novel  
CC labelled conjugate comprises at least one label attached to a functional  
CC surrogate of an analyte of interest. The surrogate is capable of  
CC competing effectively with the analyte for a limiting amount of an  
CC affinity receptor for the analyte. The conjugate exhibits an activity  
CC that is altered upon interaction with the affinity receptor and this  
CC activity can be measured and related to the amount of the analyte present  
CC in a sample. Functional surrogates such as this have an immunoreactive  
CC group that allows the surrogate to compete effectively and with the  
CC analyte for a limiting amount of its affinity receptor. Functional  
CC surrogates are able to mimic naturally occurring analytes. They can  
CC be labeled for use in standard competitive affinity assays (esp.  
CC homogeneous immunoassays) for detecting large macromolecules such as  
CC polypeptides, polysaccharides, polynucleotides, glycoproteins and  
CC lipid-containing macromolecules, as well as small haptens. Typical  
CC diagnostic analytes for detection include cardiac or tumour markers,  
CC allergens, hormones related to fertility-pregnancy or analytes  
CC associated with infectious disease. In particular, the assays are  
CC useful for detecting ferritin, follicle stimulating hormone, human  
CC growth hormone, immunoglobulin E, prolactin, parathyroid hormone, human  
CC placental lactogen, hepatitis antigens or antibodies against them, human  
CC chorionic gonadotropin, human interleukinising hormone, cytomegalovirus,  
CC Chlamydia, Streptococcus a, rubella, toxoplasma, herpes virus, DK-MB,  
CC myoglobin, myosin light chain, tropoin, carcinoembryonic antigen,  
CC alpha-fetoprotein, prostate-specific antigen and CA125 (a tumour marker).  
CC Sequence 9 AA;  
SQ

Query Match 72.0%; Score 36; DB 1; Length 9;  
Best Local Similarity 62.5%; Pred. No. 1.3e+05;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 CGRRVFC 8  
111:111  
DB 2 CGGASLFC 9

RESULT 3  
R13875  
ID R13875 standard; Protein: 248 AA.  
AC R13875:  
DT 11-OCT-1991 (first entry)  
DE Murine Cytotoxic Cell Protease-2.  
KW mouse; CCP2 inhibitor; cytotoxic T-lymphocytes; ss.  
OS Mus musculus.  
PN W09110685-A.  
PD 25-JUN-1991.  
PF 17-JAN-1991; U00340.  
PR 19-JAN-1990; US-467880.  
PA (SEBA-) SERAGEN INC.  
PI Blackley RC, Lobe CG, Paetkau VH, James MN, Murphy M;  
DR N-PSDB: Q12863.  
PT DNA vectors, and inhibitors of cytotoxic cell protease - for  
PT treatment of auto-immune diseases e.g. pernicious anaemia,  
PT rheumatoid arthritis, allo-graft rejection etc.  
PS Claim 6; Fig 10; 62pp; English.  
CC The CCP2 coding sequence was isolated from the cytotoxic T-cell  
CC line MTL 2.8.2 generated from CBA/J mice. The amino acid sequence  
CC of CCP2 was predicted from the cDNA sequence. The structure of the  
CC protein was used to design peptides which competitively inhibit the

CC protease. See Q12862-6 and R13254-R13262.  
SQ Sequence 248 AA;  
SQ

Query Match 70.0%; Score 35; DB 1; Length 248;  
Best Local Similarity 71.4%; Pred. No. 72;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGRVFC 8  
111:111  
DB 44 GGRVFC 50

RESULT 4  
W10886  
ID W10886 standard; peptide: 10 AA.  
AC W10886:  
DT 07-NOV-1997 (first entry)  
DE MAb anti-HBsAg binder sequence, M3, from R8C library.  
KW Functional surrogate; analyte; affinity receptor; immunoreactive group;  
KW mimic: homogeneous immunoassay; detection; diagnostic analyte; Chlamydia;  
KW cardiac marker; tumour marker; allergen; hormone; fertility; myoglobin;  
KW pregnancy; infectious disease; ferritin; myosin light chain; tropoin;  
KW follicle stimulating hormone; human; growth hormone; immunoglobulin E;  
KW prolactin; parathyroid hormone; placental lactogen; hepatitis antigen;  
KW antibody; chorionic gonadotropin; interleukinising hormone; cytomegalovirus;  
KW Streptococcus; rubella; toxoplasma; interleukinising hormone; cytomegalovirus;  
KW carcinoembryonic antigen; alpha-fetoprotein; herpes virus; CA125.  
OS Synthetic.  
PN W09641172-A1.  
PD 19-DEC-1996.  
PF 07-JUN-1996; U10498.  
PR 07-JUN-1995; US-476375.  
PA (CYTO-) CYTOGEN CORP.  
PI Carter JM, Lee-Own FV;  
DR WPI: 97-077284/07.  
DR N-PSDB: T48742.  
PT Labelled functional surrogate of an analyte - useful as competitor  
PT molecule in affinity assays, esp. for detecting large macromolecules  
PS such as ferritin  
PS Claim 56; Page 102; 156pp; English.  
CC This sequence represents a monoclonal anti-hepatitis B antigen binder  
CC sequence from the R8C library which may be used in the conjugate of the  
CC invention. The novel labelled conjugate comprises at least one label  
CC attached to a functional surrogate of an analyte of interest. The  
CC surrogate is capable of competing effectively with the analyte for a  
CC limiting amount of an affinity receptor for the analyte. The conjugate  
CC exhibits an activity that is altered upon interaction with the affinity  
CC receptor and this activity can be measured and related to the amount of  
CC the analyte present in a sample. Functional surrogates such as this have  
CC an immunoreactive group that allows the surrogate to compete effectively  
CC with the analyte for a limiting amount of its affinity receptor.  
CC Functional surrogates are able to mimic naturally occurring analytes.  
CC They can be labeled for use in standard competitive affinity assays  
CC (esp. homogeneous immunoassays) for detecting large macromolecules such  
CC as polypeptides, polysaccharides, polynucleotides, glycoproteins and  
CC lipid-containing macromolecules, as well as small haptens. Typical  
CC diagnostic analytes for detection include cardiac or tumour markers,  
CC allergens, hormones related to fertility-pregnancy or analytes associated  
CC with infectious disease. In particular, the assays are useful for  
CC detecting ferritin, follicle stimulating hormone, human growth hormone,  
CC immunoglobulin E, prolactin, parathyroid hormone, human placental  
CC lactogen, hepatitis antigens or antibodies against them, human  
CC chorionic gonadotropin, human interleukinising hormone, cytomegalovirus,  
CC Chlamydia, Streptococcus a, rubella, toxoplasma, herpes virus, DK-MB,  
CC myoglobin, myosin light chain, tropoin, carcinoembryonic antigen,  
CC alpha-fetoprotein, prostate-specific antigen and CA125 (a tumour marker).  
CC Sequence 10 AA;  
SQ

Query Match 70.0%; Score 35; DB 1; Length 10;  
Best Local Similarity 71.4%; Pred. No. 4.6;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGGRKVF 7  
 111111  
 Db 1 CGGRSIF 7

RESULT 5  
 P92219  
 ID P92219 standard; protein; 2317 AA.  
 AC P92219;  
 DT 22-FEB-1990 (first entry)  
 DE CRI protein  
 KW Complement; cofactor.  
 OS Homo sapiens (human).  
 FH Key Location/Qualifiers  
 FT peptide 10..50  
 FT /label= signal\_peptide  
 PN W08999220-A.  
 PD 05-OCT-1989.  
 PE 31-MAR-1989; U01358.  
 PR 01-APR-1988; US-176532.  
 PA (TCEL) T Cell Sciences Inc; (UYUO) The Johns Hopkins University;  
 PA (BRIG\*) The Brigham and Women's Hospital.  
 PI Reardon DT, Klickestein LB, Wong W, Carson G, Concino MF, Makrides SC;  
 DR WPI: 89-309498/42.  
 DR N-PSDB: N91477.  
 PT New nucleic acid sequences encoding new CRI protein - and its fragment,  
 PT for diagnosis and control of complement-related immune defects,  
 PS inflammation, myocardial infarct, etc  
 PS Claim 1, Fig. 1; 191PP; English.  
 CC This is full-length CRI protein, and shortened forms are new, lacking  
 CC the transmembrane region. The proteins and fragments bind C3b and/or  
 CC C4b, have cofactor activity and inhibit C3 and C5 convertase activity.  
 CC In the sequence, x-untranslated region. This has 7 short consensus  
 CC repeats (CR) of 4 long direct homologous repeats (LHR) (see fig. 10 in  
 CC patent). Deletion mutants, eg LHR-A for binding of C4b, and LHR-C  
 CC for C3b binding, be made, and LHR-B and -C can detect CR2 sequences.  
 CC They are useful in diagnosing and treating immune disorders, and prevent  
 CC perfusion injury.  
 CC Sequence 2317 AA;  
 SQ

Query Match 64.0%; Score 32; DB 1; Length 2317;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGRKVF 7  
 111111  
 Db 209 GGRKVF 214

RESULT 6  
 R06052  
 ID R06052 standard; protein; 18 AA.  
 AC R06052;  
 DT 20-NOV-1990 (first entry)  
 DE Immunoreactive peptide 4S36, antigenic to the gp41 peptide of HIV-1.  
 KW HIV-1; HIV-2; AIDS; gp41; gp32; vaccine.  
 OS HIV.  
 PN W09007119-A.  
 PD 28-JUN-1990.  
 PE 15-DEC-1989; U05640.  
 PR 20-DEC-1988; US-287412.  
 PA (IMMU-) IMMUNODIAGNOSTICS I.  
 PI FORMOSO C, OLSEN DA, BUCHANAN TM;  
 DR WPI: 90-224611/29.  
 PT Synthetic HIV-like peptide(s) - contg. immuno-reactive regions of  
 PT gp41 of HIV-1 or gp32 of HIV-2 used for detecting antibodies or  
 PT producing antibodies.  
 PS Claim 2; Page 37; 63pp; English.  
 CC Peptides may be chemically linked to a carrier protein giving  
 CC a more accurate test for HIV-1 and HIV-2 with a greater definition  
 CC between positive and negative results. The may also be used to

CC generate an immune response in primates.  
 SQ Sequence 18 AA;

Query Match 64.0%; Score 32; DB 1; Length 18;  
 Best Local Similarity 50.0%; Pred. No. 23;  
 Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGGRKVF 8  
 111111  
 Db 4 CGGRKVF 11

RESULT 7  
 R11810  
 ID R11810 standard; protein; 2039 AA.  
 AC R11810;  
 DT 25-JUN-1991 (first entry)  
 DE Human complement type 1 receptor.  
 KW Complement system; C3b/C4b receptor; CRI; allergic reaction;  
 KW immune response; clone lambda T109.1.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT peptide 1..41  
 FT /label= putative signal peptide  
 FT protein 42..2039  
 FT /label= CRI  
 PN W09105047-A.  
 PD 18-APR-1991.  
 PE 25-SEP-1990; U05454.  
 PR 26-SEP-1989; US-412745.  
 PR 26-SEP-1990; US-912349.  
 PA (TCEL-) T CELL SCI INC.  
 PA (UYUO) JOHNS HOPKINS UNIVERSITY.  
 PA (BRIG-) BRIGHAM AND WOMEN'S HOSPITAL.  
 PI Reardon DT, Klickestein LB, Wong WW, Carson GR, Hoh M, Concino MF,  
 PI Makrides SC, Marsh HC;  
 DR WPI: 91-132854/18.  
 DR N-PSDB: Q11642.  
 PT Human complement receptor type 1 gene, encoded proteins and  
 PT fragments - for treatment of immune disorders, myocardial infarct,  
 PT damage due to inflammation and in treatment of thrombosis  
 PS Claim 41; Fig 1; 234PP; English.  
 CC The invention also covers fragments of this protein which have the  
 CC ability to bind C3b and/or C4b, have cofactor I activity or can  
 CC inhibit C3 or C5 convertase activity. The full-length protein, or  
 CC its specified fragments are used to treat patients with immune  
 CC disorders or a disorder caused by inappropriate complement  
 CC activity. The protein is also used to treat thrombotic conditions  
 CC in humans and animals. See also Q11643.  
 CC Sequence 2039 AA;  
 SQ

Query Match 64.0%; Score 32; DB 1; Length 2039;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+03;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGRKVF 7  
 111111  
 Db 200 GGRKVF 205

RESULT 8  
 R11982  
 ID R11982 standard; protein; 1537 AA.  
 AC R11982;  
 DT 25-JUN-1991 (first entry)  
 DE Partial human complement type 1 receptor.  
 KW Complement system; C3b/C4b receptor; CRI; allergic reaction;  
 KW immune response; long homologous repeat; LHR.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT region 1..438

FT                   /label= LHR-B  
FT                   439..891  
FT                   /label= LHR-C  
FT                   892..1341  
FT                   /label= LHR-D  
FT                   1495..1498  
FT                   /note= "positively-charged; preceded by hydrophobic  
FT                   sequence"  
FT                   1521..1526  
FT                   /note= "has 67 per cent homology to site of protein  
FT                   kinase C phosphorylation in the EGF  
FT                   receptor"  
FT                   region  
FT                   WO9105047-A.  
PN                   18-APR-1991.  
PD                   18-APR-1991.  
PF                   25-SEP-1990: U05454.  
PR                   26-SEP-1989: US-412745.  
PR                   26-SEP-1990: US-912349.  
PA                   (TCELL-) T CELL SCI INC.  
PA                   (UNJO ) JOHNS HOPKINS UNIVERSITY.  
PA                   (BRIG-) BRIGHAM AND WOMEN'S HOSPITAL.  
PI                   Fearon Dr, Klicstein LB, Wong WM, Carson GR, Hoh M, Concino MF,  
PI                   Makrides SC, Marsh HC;  
DR                   WPI: 91-132854/18.  
DR                   N-PSDB: Q11643.  
PT                   Human complement receptor type 1 gene, encoded proteins and  
PT                   fragments - for treatment of immune disorders, myocardial infarct,  
PT                   damage due to inflammation and in treatment of thrombosis  
PS                   Disclosure: Fig 5: 234pp: English.  
CC                   This sequence comprises three of the four tandem, direct, long  
CC                   homologous repeats of the full-length F allotype of CRI. LHR-A is  
CC                   absent. Each LHR might represent a single C3b/C4b binding domain,  
CC                   making the receptor multivalent. The LHR's are composed of 7 short  
CC                   consensus repeats of 60-70 residues resembling the SCR's of other  
CC                   C3/C4 binding proteins. The protein and fragments of it having C3b  
CC                   and/or C4b binding activity can be used to treat immune disorders  
CC                   or disorders involving inappropriate complement activity.  
CC                   See also Q11642.  
SQ                   Sequence   1537 AA:

Query Match                   64.0%; Score 32; DB 1; Length 1537;  
Best Local Similarity   100.0%; Pred. No. 1e+03;  
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

OY                   2 GGRKVF 7  
                  |||||  
DB                   148 GGRKVF 153

RESULT   9  
ID   R28543  
AC   R28543: standard: peptide; 543 AA.  
DT   19-MAR-1993 (first entry)  
DE   CRI-4 (35E, 37Y) analogue.  
KW   short consensus repeat; regulator of complement activation;  
KM   C3b binding; C4b binding; human complement type 1 receptor.  
OS   Homo sapiens.  
FH   Key                   Location/Qualifiers  
FT                   1..60  
FT                   /label= SCR-1  
FT                   61..122  
FT                   /label= SCR-2  
FT                   451..510  
FT                   /label= SCR-8  
FT                   511..543  
FT                   /label= SCR-9  
FT                   /note= "TRUNCATED"  
FT                   misc-difference 35  
FT                   /note= "Gly substituted by Glu (SCR-8)"  
FT                   misc-difference 37  
FT                   /note= "Ser substituted by Tyr (SCR-8)"  
PN   EP-512733-A.

PD   11-NOV-1992.  
PF   28-APR-1992: 303826.  
PR   03-MAY-1991: US-695514.  
PA   (UNITV ) UNIV WASHINGTON.  
PI   Atkinson JP, Hourcade D, Kiyoh M;  
DR   WPI: 92-375009/46.  
PT   Complement activity regulator protein analogues - useful for  
PT   treating auto-immune diseases, to suppress transplant rejection,  
PT   for diagnosis etc.  
PS   Claim 11: Fig 2 and R11810: 23pp: English.  
CC   The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)  
CC   168:1255-1270. It encodes the first 8 and a half amino terminal  
CC   SCRs of CRI. The invention concerns analogues of "regulator of  
CC   complement activation" proteins or truncated, hybrid or recombinant  
CC   forms of them. CRI-4 is a preferred truncated form and a number of  
CC   specific substitution variants of it are claimed. Positions 35 and  
CC   37 of SCR-1 and the corresponding positions in SCR-8 have been  
CC   identified as important in C4b binding. The specification does not  
CC   contain the CRI-4 sequence; the sequence given here was constructed  
CC   from the full-length CRI amino acid sequence having GENESEQ  
SQ   accession number R11810 and descriptions in the disclosure.  
SQ   Sequence   543 AA:

Query Match                   64.0%; Score 32; DB 1; Length 543;  
Best Local Similarity   100.0%; Pred. No. 4.2e+02;  
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

OY                   2 GGRKVF 7  
                  |||||  
DB                   159 GGRKVF 164

RESULT   10  
ID   R28544  
AC   R28544: standard: peptide; 543 AA.  
DT   19-MAR-1993 (first entry)  
DE   CRI-4 (35E) analogue.  
KW   short consensus repeat; regulator of complement activation;  
KM   C3b binding; C4b binding; human complement type 1 receptor.  
OS   Homo sapiens.  
FH   Key                   Location/Qualifiers  
FT                   1..60  
FT                   /label= SCR-1  
FT                   61..122  
FT                   /label= SCR-2  
FT                   451..510  
FT                   /label= SCR-8  
FT                   511..543  
FT                   /label= SCR-9  
FT                   /note= "TRUNCATED"  
FT                   misc-difference 35  
FT                   /note= "Gly substituted by Glu (SCR-8)"  
FT                   EP-512733-A.  
PN   11-NOV-1992.  
PF   28-APR-1992: 303826.  
PR   03-MAY-1991: US-695514.  
PA   (UNITV ) UNIV WASHINGTON.  
PI   Atkinson JP, Hourcade D, Kiyoh M;  
DR   WPI: 92-375009/46.  
PT   Complement activity regulator protein analogues - useful for  
PT   treating auto-immune diseases, to suppress transplant rejection,  
PT   for diagnosis etc.  
PS   Claim 11: Fig 2 and R11810: 23pp: English.  
CC   The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)  
CC   168:1255-1270. It encodes the first 8 and a half amino terminal  
CC   SCRs of CRI. The invention concerns analogues of "regulator of  
CC   complement activation" proteins or truncated, hybrid or recombinant  
CC   forms of them. CRI-4 is a preferred truncated form and a number of  
CC   specific substitution variants of it are claimed. Positions 35 and  
CC   37 of SCR-1 and the corresponding positions in SCR-8 have been  
CC   identified as important in C4b binding. The specification does not

CC contain the CRI-4 sequence; the sequence given here was constructed  
 CC from the full-length CRI amino acid sequence having GENSEQ  
 CC accession number R1810 and descriptions in the disclosure.  
 SQ Sequence 543 AA;

Query Match 64.0%; Score 32; DB 1; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGRKVF 7  
 |||||  
 DB 159 GGRKVF 164

RESULT 11  
 R28545  
 ID R28545 standard; peptide; 543 AA.  
 AC R28545;  
 DT 19-MAR-1993 (first entry)  
 DE CRI-4 (37Y) analogue.  
 KW Short consensus repeat; regulator of complement activation;  
 KW C3b binding; C4b binding; human complement type 1 receptor.  
 OS Homo sapiens.

FH Key Location/Qualifiers  
 FT region 1. .60  
 FT /label= SCR-1  
 FT region 61. .122  
 FT /label= SCR-2  
 FT region 451. .510  
 FT /label= SCR-8  
 FT region 511. .543  
 FT /label= SCR-9  
 FT /note= "TRUNCATED"

FT misc\_difference 37  
 FT /note= "Ser substituted by Tyr (SCR-8)"

PN EP-512733-A.  
 PD 11-NOV-1992.  
 PF 28-APR-1992; 303826.  
 PR 03-MAY-1991; US-695514.  
 PA (UNIV ) UNIV WASHINGTON.  
 PI Atkinson JP, Hourcade D, Kirsch M;  
 DR WPI; 92-375009/46.  
 PT Complement activity regulator protein analogues - useful for  
 PT treating auto-immune diseases, to suppress transplant rejection,  
 PT for diagnosis etc.  
 PS Claim 11: Fig 2 and R1810; 23pp; English.  
 CC The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)  
 CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
 CC SCRs of CRI. The invention concerns analogues of "regulator of  
 CC complement activation" proteins or truncated, hybrid or recombinant  
 CC forms of them. CRI-4 is a preferred truncated form and a number of  
 CC specified substitution variants of it are claimed. Positions 35 and  
 CC 37 of SCR-1 and the corresponding positions in SCR-8 have been  
 CC identified as important in C4b binding. The specification does not  
 CC contain the CRI-4 sequence; the sequence given here was constructed  
 CC from the full-length CRI amino acid sequence having GENSEQ  
 CC accession number R1810 and descriptions in the disclosure.  
 SQ Sequence 543 AA;

Query Match 64.0%; Score 32; DB 1; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGRKVF 7  
 |||||  
 DB 159 GGRKVF 164

RESULT 12  
 R28546  
 ID R28546 standard; peptide; 543 AA.

AC R28546;  
 DT 19-MAR-1993 (first entry)  
 DE CRI-4 (44T, 47D, 49L) analogue.  
 KW Short consensus repeat; regulator of complement activation;  
 KW C3b binding; C4b binding; human complement type 1 receptor.  
 OS Homo sapiens.

FH Key Location/Qualifiers  
 FT region 1. .60  
 FT /label= SCR-1  
 FT region 61. .122  
 FT /label= SCR-2  
 FT region 451. .510  
 FT /label= SCR-8  
 FT region 511. .543  
 FT /label= SCR-9  
 FT /note= "TRUNCATED"

FT misc\_difference 44  
 FT /note= "Ile substituted by Thr (SCR-8)"  
 FT misc\_difference 47  
 FT /note= "Lys substituted by Asp (SCR-8)"  
 FT misc\_difference 49  
 FT /note= "Ser substituted by Leu (SCR-8)"

PN EP-512733-A.  
 PD 11-NOV-1992.  
 PF 28-APR-1992; 303826.  
 PR 03-MAY-1991; US-695514.  
 PA (UNIV ) UNIV WASHINGTON.  
 PI Atkinson JP, Hourcade D, Kirsch M;  
 DR WPI; 92-375009/46.  
 PT Complement activity regulator protein analogues - useful for  
 PT treating auto-immune diseases, to suppress transplant rejection,  
 PT for diagnosis etc.  
 PS Claim 11: Fig 2 and R1810; 23pp; English.  
 CC The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)  
 CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
 CC SCRs of CRI. The invention concerns analogues of "regulator of  
 CC complement activation" proteins or truncated, hybrid or recombinant  
 CC forms of them. CRI-4 is a preferred truncated form and a number of  
 CC specified substitution variants of it are claimed in which certain  
 CC positions in SCR-1 which have been identified as important for the  
 CC degree of C3b- and C4b-binding are substituted by amino acids from  
 CC the corresponding positions in SCR-8. The specification does not  
 CC contain the CRI-4 sequence; the sequence given here was constructed  
 CC from the full-length CRI amino acid sequence having GENSEQ  
 CC accession number R1810 and descriptions in the disclosure.  
 SQ Sequence 543 AA;

Query Match 64.0%; Score 32; DB 1; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGRKVF 7  
 |||||  
 DB 159 GGRKVF 164

RESULT 13  
 R28547  
 ID R28547 standard; peptide; 543 AA.  
 AC R28547;  
 DT 19-MAR-1993 (first entry)  
 DE CRI-4 (52S, 53S, 54P) analogue.  
 KW Short consensus repeat; regulator of complement activation;  
 KW C3b binding; C4b binding; human complement type 1 receptor.  
 OS Homo sapiens.

FH Key Location/Qualifiers  
 FT region 1. .60  
 FT /label= SCR-1  
 FT region 61. .122  
 FT /label= SCR-2  
 FT region 451. .510  
 FT /label= SCR-8

FT region 511..543  
FT /label= SCR-9  
FT /note= "TRUNCATED"  
FT misc-difference 52  
FT /note= "Thr substituted by Ser (SCR-8)"  
FT misc-difference 53  
FT /note= "Gly substituted by Ser (SCR-8)"  
FT misc-difference 54  
FT /note= "Ala substituted by Pro (SCR-8)"  
PN EP-512733-A.  
PD 11-NOV-1992.  
PD 28-APR-1992; 303826.  
PR 03-MAY-1991; US-695514.  
PA (UNIM ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Kiyeh M;  
DR WPI; 92-375009/46.  
PT Complement activity regulator protein analogues - useful for  
PT treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11; Fig 2 and R11810; 23pp; English.  
CC The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCRs of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-1 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-8. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQ  
CC accession number R11810 and descriptions in the disclosure.  
SQ Sequence 543 AA;

Query Match 64.0%; Score 32; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGRKVF 7  
|11111|  
DB 159 GGRKVF 164

RESULT 14  
R28548  
ID R28548 standard; peptide: 543 AA.  
AC R28548; 19-MAR-1993 (first entry)  
DE CRI-4 (57V, 59K) analogue.  
KW short consensus repeat; regulator of complement activation;  
KW C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT region 1..60  
FT /label= SCR-1  
FT 61..122  
FT /label= SCR-2  
FT 451..510  
FT /label= SCR-8  
FT 511..543  
FT /label= SCR-9  
FT /note= "TRUNCATED"  
FT misc-difference 57  
FT /note= "Arg substituted by Val (SCR-8)"  
FT misc-difference 59  
FT /note= "Arg substituted by Lys (SCR-8)"  
PN EP-512733-A.  
PD 11-NOV-1992.  
PD 28-APR-1992; 303826.  
PR 03-MAY-1991; US-695514.  
PA (UNIM ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Kiyeh M;

DR WPI; 92-375009/46.  
PT Complement activity regulator protein analogues - useful for  
PT treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11; Fig 2 and R11810; 23pp; English.  
CC The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCRs of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-1 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-8. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQ  
CC accession number R11810 and descriptions in the disclosure.  
SQ Sequence 543 AA;

Query Match 64.0%; Score 32; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGRKVF 7  
|11111|  
DB 159 GGRKVF 164

RESULT 15  
R28549  
ID R28549 standard; peptide: 543 AA.  
AC R28549;  
DE 19-MAR-1993 (first entry)  
DE CRI-4 (64K, 65T) analogue.  
KW short consensus repeat; regulator of complement activation;  
KW C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT region 1..60  
FT /label= SCR-1  
FT 61..122  
FT /label= SCR-2  
FT 451..510  
FT /label= SCR-8  
FT 511..543  
FT /label= SCR-9  
FT /note= "TRUNCATED"  
FT misc-difference 64  
FT /note= "Arg substituted by Lys (SCR-9)"  
FT misc-difference 65  
FT /note= "Asn substituted by Thr (SCR-9)"  
PN EP-512733-A.  
PD 11-NOV-1992.  
PD 28-APR-1992; 303826.  
PR 03-MAY-1991; US-695514.  
PA (UNIM ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Kiyeh M;  
DR WPI; 92-375009/46.  
PT Complement activity regulator protein analogues - useful for  
PT treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11; Fig 2 and R11810; 23pp; English.  
CC The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCRs of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-2 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-9. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed



CC from the full-length CRI amino acid sequence having GENESEQ  
 CC accession number R11810 and descriptions in the disclosure.  
 SQ Sequence 543 AA;

Query Match 64.0%; Score 32; DB 1; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGRKVF 7  
 |||||  
 Db 159 GGRKVF 164

Search completed: July 18, 1999, 00:42:55  
 Job time: 8588 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 06:07:57 ; Search time 39.54 seconds  
(without alignments)  
1.997 Million cell updates/sec

Title: US-09-142-043-7  
Perfect score: 50  
Sequence: 1 CGGRKVF 8

Scoring table: BLOSUM62

Searched: 106577 seqs, 9868381 residues

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/prodata/1/iaa/5A.COMB.pep:\*  
2: /cgn2\_6/prodata/1/iaa/5B.COMB.pep:\*  
3: /cgn2\_6/prodata/1/iaa/PCTUS9.COMB.pep:\*  
4: /cgn2\_6/prodata/1/iaa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	32	64.0	197	2	US-08-356-361-27	Sequence 27, Appl
2	32	64.0	76	2	US-08-356-361-28	Sequence 28, Appl
3	32	64.0	254	2	US-08-356-361-29	Sequence 29, Appl
4	32	64.0	254	2	US-08-356-361-30	Sequence 30, Appl
5	32	64.0	133	2	US-08-356-361-31	Sequence 31, Appl
6	32	64.0	1025	2	US-08-304-309-2	Sequence 2, Appl
7	32	64.0	1025	2	US-08-304-309-4	Sequence 2, Appl
8	32	64.0	197	2	US-08-769-967A-27	Sequence 27, Appl
9	32	64.0	76	2	US-08-769-967A-28	Sequence 28, Appl
10	32	64.0	254	2	US-08-769-967A-29	Sequence 29, Appl
11	32	64.0	254	2	US-08-769-967A-30	Sequence 30, Appl
12	32	64.0	133	2	US-08-769-967A-31	Sequence 31, Appl
13	32	64.0	1025	2	US-08-304-309-2	Sequence 2, Appl
14	32	64.0	1025	2	US-08-304-309-4	Sequence 2, Appl
15	32	64.0	1025	2	US-08-304-309-4	Sequence 2, Appl
16	32	64.0	18	1	US-08-182-483A-17	Sequence 17, Appl
17	32	64.0	18	1	US-08-182-483A-19	Sequence 19, Appl
18	32	64.0	18	1	US-08-243-879A-16	Sequence 16, Appl
19	32	64.0	18	1	US-08-243-879A-18	Sequence 18, Appl
20	32	64.0	18	1	US-08-229-418-2	Sequence 2, Appl
21	32	64.0	18	2	US-08-499-523-37	Sequence 37, Appl
22	32	64.0	18	2	US-08-499-523-39	Sequence 39, Appl
23	32	64.0	353	2	PCT-US95-04464-2	Sequence 2, Appl
24	32	64.0	8	2	US-08-397-633A-100	Sequence 100, App
25	32	64.0	856	1	US-08-022-835-2	Sequence 2, Appl
26	32	64.0	856	1	US-08-022-835-4	Sequence 4, Appl
27	32	64.0	855	1	US-08-022-835-6	Sequence 6, Appl
28	32	64.0	855	1	US-08-127-499A-14	Sequence 14, Appl
29	32	64.0	20	1	US-08-218-025A-64	Sequence 64, Appl
30	32	64.0	861	1	US-08-482-847-14	Sequence 14, Appl
31	32	64.0	856	1	US-08-388-809-2	Sequence 2, Appl
32	32	64.0	857	1	US-08-388-809-4	Sequence 4, Appl
33	32	64.0	855	1	US-08-388-809-6	Sequence 6, Appl
34	32	64.0	826	1	US-08-375-510-2	Sequence 2, Appl
35	32	64.0	519	1	US-08-589-446-8	Sequence 8, Appl
36	32	64.0	519	1	US-08-444-882-8	Sequence 8, Appl
37	32	64.0	856	2	US-08-375-100-1	Sequence 27, Appl
38	32	64.0	620	2	US-08-706-037-27	Sequence 27, Appl
39	32	64.0	74	2	US-08-940-661A-2	Sequence 2, Appl
					US-08-379-538-2	Sequence 2, Appl

ALIGNMENTS

40	29	58.0	519	2	US-08-389-459A-8	Sequence 8, Appl
41	29	58.0	826	2	US-08-487-657-2	Sequence 2, Appl
42	29	58.0	23	2	US-08-493-235-29	Sequence 29, Appl
43	29	58.0	23	2	US-08-493-235-30	Sequence 30, Appl
44	29	58.0	880	2	US-08-788-815-7	Sequence 7, Appl
45	29	58.0	551	2	US-08-417-210A-137	Sequence 137, App

RESULT 1  
US-08-356-361-27  
; Sequence 27, Application US/08356361  
; Patent No. 5833989  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Richard A.G.  
; APPLICANT: Dodd, Ian  
; APPLICANT: Freeman Mary A.  
; APPLICANT: Mossakowska, Danuta E.I.  
; TITLE OF INVENTION: No. 5833989e1 Compounds  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSER: SmithKline Beecham-Corporate Intellectual Property  
; STREET: P.O. Box 1539  
; CITY: King of Prussia  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19406  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentia Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/356,361  
; FILING DATE: 03-Jul-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jervais, Herbert H.  
; REGISTRATION NUMBER: 31,171  
; REFERENCE/DOCKET NUMBER: P30423  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (610) 270-5019  
; TELEFAX: (610) 270-5090  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 197 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: N-terminal  
; US-08-356-361-27

Query Match 64.0%; Score 32; DB 2; Length 197;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGRKVF 7  
DB 160 GGRKVF 165

RESULT 2  
US-08-356-361-28  
; Sequence 28, Application US/08356361  
; Patent No. 5833989  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Richard A.G.  
; APPLICANT: Dodd, Ian  
; APPLICANT: Freeman Mary A.  
; APPLICANT: Mossakowska, Danuta E.I.

TITLE OF INVENTION: No. 5833989e1 Compounds  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham-Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,361  
FILING DATE: 03-Jul-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jervis, Herbert H.  
REGISTRATION NUMBER: 31,171  
REFERENCE/DOCKET NUMBER: P30423  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5019  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 76 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-356-361-28

Query Match 64.0%; Score 32; DB 2; Length 76;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGRKVF 7  
|||||  
DB 39 GGRKVF 44

RESULT 3  
US-08-356-361-29  
Sequence 29, Application US/08356361  
Patent No. 5833989  
GENERAL INFORMATION:  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman Mary A.  
APPLICANT: Mossakowska, Danuta E.I.  
TITLE OF INVENTION: No. 5833989e1 Compounds  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham-Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,361  
FILING DATE: 03-Jul-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jervis, Herbert H.

REGISTRATION NUMBER: 31,171  
REFERENCE/DOCKET NUMBER: P30423  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5019  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-356-361-29

Query Match 64.0%; Score 32; DB 2; Length 254;  
Best Local Similarity 100.0%; Pred. No. 11e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGRKVF 7  
|||||  
DB 160 GGRKVF 165

RESULT 4  
US-08-356-361-30  
Sequence 30, Application US/08356361  
Patent No. 5833989  
GENERAL INFORMATION:  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman Mary A.  
APPLICANT: Mossakowska, Danuta E.I.  
TITLE OF INVENTION: No. 5833989e1 Compounds  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham-Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,361  
FILING DATE: 03-Jul-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jervis, Herbert H.  
REGISTRATION NUMBER: 31,171  
REFERENCE/DOCKET NUMBER: P30423  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5019  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-356-361-30

Query Match 64.0%; Score 32; DB 2; Length 254;  
Best Local Similarity 100.0%; Pred. No. 11e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGRKVF 7

Db 160 GGRVVF 165

RESULT 5  
US-08-356-361-31  
Sequence 31, Application US/08356361

Patent No. 5833989

GENERAL INFORMATION:

APPLICANT: Smith, Richard A.G.

APPLICANT: Dodd, Ian

APPLICANT: Freeman Mary A.

APPLICANT: Mossakowska, Danuta E.I.

TITLE OF INVENTION: No. 5833989e1 Compounds

NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham-Corporate Intellectual Property

STREET: P.O. Box 1539

CITY: King of Prussia

STATE: Pennsylvania

COUNTRY: USA

ZIP: 19406

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/356,361

FILING DATE: 03-Jul-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Jervis, Herbert H.

REGISTRATION NUMBER: 31,171

REFERENCE/DOCKET NUMBER: P30423

TELECOMMUNICATION INFORMATION:

TELEPHONE: (610) 270-5019

TELEFAX: (610) 270-5090

INFORMATION FOR SEQ ID NO: 31:

SEQUENCE CHARACTERISTICS:

LENGTH: 133 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

FRAGMENT TYPE: internal

US-08-356-361-31

Query Match 64.0%; Score 32; DB 2; Length 133;

Best Local Similarity 100.0%; Pred. No. 61;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGRVVF 7

Db 39 GGRVVF 44

RESULT 6  
US-08-304-309-2  
Sequence 2, Application US/08304309

Patent No. 5856454

GENERAL INFORMATION:

APPLICANT: GONZALEZ, Frank J.

APPLICANT: FERNANDEZ-SALGUERO, Pedro

TITLE OF INVENTION: CLONING AND EXPRESSION OF CDNA FOR HUMAN

TITLE OF INVENTION: DIHYDROPYRIMIDINE DEHYDROGENASE

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend Kourie and Crew

STREET: Stewart Street Tower, One Market Plaza

CITY: San Francisco

STATE: California

COUNTRY: US

ZIP: 94105-1493

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/304,309

FILING DATE: 09-SEP-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Smith, Timothy L.

REGISTRATION NUMBER: 35,367

REFERENCE/DOCKET NUMBER: 15280-210

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 543-9600

TELEFAX: (415) 543-5043

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1025 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-304-309-2

Query Match 64.0%; Score 32; DB 2; Length 1025;

Best Local Similarity 71.4%; Pred. No. 3,5e+02;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGGRVVF 7

Db 354 CGARVVF 360

RESULT 7

US-08-304-309-4

Sequence 4, Application US/08304309

Patent No. 5856454

GENERAL INFORMATION:

APPLICANT: GONZALEZ, Frank J.

APPLICANT: FERNANDEZ-SALGUERO, Pedro

TITLE OF INVENTION: CLONING AND EXPRESSION OF CDNA FOR HUMAN

TITLE OF INVENTION: DIHYDROPYRIMIDINE DEHYDROGENASE

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend Kourie and Crew

STREET: Stewart Street Tower, One Market Plaza

CITY: San Francisco

STATE: California

COUNTRY: US

ZIP: 94105-1493

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/304,309

FILING DATE: 09-SEP-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Smith, Timothy L.

REGISTRATION NUMBER: 35,367

REFERENCE/DOCKET NUMBER: 15280-210

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 543-9600

TELEFAX: (415) 543-5043

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1025 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-304-309-4

Query Match 64.0%; Score 32; DB 2; Length 1025;  
Best Local Similarity 71.4%; Pred. No. 3.5e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGRKVF 7  
111111  
DB 354 GGRKVF 360

RESULT 8  
US-08-769-967A-27  
Sequence 27, Application US/08769967A  
Patent No. 5859223

GENERAL INFORMATION:

APPLICANT: Mossakowska, Danuta E.I.

APPLICANT: Smith, Richard A.G.

APPLICANT: Dodd, Ian

APPLICANT: Freeman, Anne Mary

TITLE OF INVENTION: Soluble CRI Derivatives

NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporate Intellectual Property

STREET: P.O. Box 1539

CITY: King of Prussia

STATE: Pennsylvania

COUNTRY: USA

ZIP: 19406

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/769,967A

FILING DATE:

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/440,569

FILING DATE: 15-May-1995

ATTORNEY/AGENT INFORMATION:

NAME: King, William T.

REGISTRATION NUMBER: 30,954

REFERENCE/DOCKET NUMBER: P30423C2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (610) 270-5364

TELEFAX: (610) 270-5090

INFORMATION FOR SEQ ID NO: 27:

SEQUENCE CHARACTERISTICS:

LENGTH: 197 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

FRAGMENT TYPE: N-terminal

US-08-769-967A-27

Query Match 64.0%; Score 32; DB 2; Length 197;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGRKVF 7  
111111  
DB 160 GGRKVF 165

RESULT 9  
US-08-769-967A-28  
Sequence 28, Application US/08769967A  
Patent No. 5859223

GENERAL INFORMATION:

APPLICANT: Mossakowska, Danuta E.I.

APPLICANT: Smith, Richard A.G.

APPLICANT: Dodd, Ian

APPLICANT: Freeman, Anne Mary

TITLE OF INVENTION: Soluble CRI Derivatives

NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporate Intellectual Property

STREET: P.O. Box 1539

CITY: King of Prussia

STATE: Pennsylvania

COUNTRY: USA

ZIP: 19406

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/769,967A

FILING DATE:

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/440,569

FILING DATE: 15-May-1995

ATTORNEY/AGENT INFORMATION:

NAME: King, William T.

REGISTRATION NUMBER: 30,954

REFERENCE/DOCKET NUMBER: P30423C2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (610) 270-5364

TELEFAX: (610) 270-5090

INFORMATION FOR SEQ ID NO: 28:

SEQUENCE CHARACTERISTICS:

LENGTH: 76 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

FRAGMENT TYPE: internal

US-08-769-967A-28

Query Match 64.0%; Score 32; DB 2; Length 76;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGRKVF 7  
111111  
DB 39 GGRKVF 44

RESULT 10  
US-08-769-967A-29  
Sequence 29, Application US/08769967A  
Patent No. 5859223

GENERAL INFORMATION:

APPLICANT: Mossakowska, Danuta E.I.

APPLICANT: Smith, Richard A.G.

APPLICANT: Dodd, Ian

APPLICANT: Freeman, Anne Mary

TITLE OF INVENTION: Soluble CRI Derivatives

NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporate Intellectual Property

STREET: P.O. Box 1539

CITY: King of Prussia

STATE: Pennsylvania

COUNTRY: USA

ZIP: 19406

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,967A  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,569  
FILING DATE: 15-May-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: P30423C2  
TELEPHONE: (610) 270-5364  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-769-967A-29

Query Match 64.0%; Score 32; DB 2; Length 254;  
Best Local Similarity 100.0%; Pred. No. 1,1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGRKVF 7  
|||||  
Db 160 GGRKVF 165

RESULT 11  
US-08-769-967A-30  
Sequence 30, Application US/08769967A  
Patent No. 5859223  
GENERAL INFORMATION:  
APPLICANT: Mossakowska, Danuta E.I.  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman, Anne Mary  
TITLE OF INVENTION: Soluble CRI Derivatives  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporate Intellectual Property  
STREET: P.O. Box 1539  
City: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,967A  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,569  
FILING DATE: 15-May-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: P30423C2  
TELEPHONE: (610) 270-5364  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 30:

SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-769-967A-30

Query Match 64.0%; Score 32; DB 2; Length 254;  
Best Local Similarity 100.0%; Pred. No. 1,1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGRKVF 7  
|||||  
Db 160 GGRKVF 165

RESULT 12  
US-08-769-967A-31  
Sequence 31, Application US/08769967A  
Patent No. 5859223  
GENERAL INFORMATION:  
APPLICANT: Mossakowska, Danuta E.I.  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman, Anne Mary  
TITLE OF INVENTION: Soluble CRI Derivatives  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporate Intellectual Property  
STREET: P.O. Box 1539  
City: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,967A  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,569  
FILING DATE: 15-May-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: P30423C2  
TELEPHONE: (610) 270-5364  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 133 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-769-967A-31

Query Match 64.0%; Score 32; DB 2; Length 133;  
Best Local Similarity 100.0%; Pred. No. 61;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGRKVF 7  
|||||  
Db 39 GGRKVF 44

RESULT 13  
PCT-US95-04567-2  
Sequence 2, Application PC/TUS9504567  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: DIHYDROPYRIMIDINE DEHYDROGENASE  
TITLE OF INVENTION: COMPOSITIONS AND METHODS OF USE  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: United States of America  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII  
SOFTWARE: Patentin Release #1.0, Version  
SOFTWARE: #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/04567  
FILING DATE: CONCURRENTLY HERewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/227,357  
FILING DATE: 13-APR-1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Wilson, Mark B.  
REGISTRATION NUMBER: 37,259  
REFERENCE/DOCKET NUMBER: UOAB025P--  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (713) 789-2679  
TELEX: 79-0924  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1025 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-04567-2

Query Match 64.0%; Score 32; DB 3; Length 1025;  
Best Local Similarity 71.4%; Pred. No. 3.5e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGRRVF 7  
|||1-||  
DB 354 CGRRVF 360

RESULT 14  
PCT-US95-04567-4  
Sequence 4, Application PC/TUS9504567  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: DIHYDROPYRIMIDINE DEHYDROGENASE  
TITLE OF INVENTION: COMPOSITIONS AND METHODS OF USE  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: United States of America  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII  
SOFTWARE: Patentin Release #1.0, Version  
SOFTWARE: #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/04567  
FILING DATE: CONCURRENTLY HERewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/227,357  
FILING DATE: 13-APR-1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Wilson, Mark B.  
REGISTRATION NUMBER: 37,259  
REFERENCE/DOCKET NUMBER: UOAB025P--  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (713) 789-2679  
TELEX: 79-0924  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1025 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-04567-4

Query Match 64.0%; Score 32; DB 3; Length 1025;  
Best Local Similarity 71.4%; Pred. No. 3.5e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGRRVF 7  
|||1-||  
DB 354 CGRRVF 360

RESULT 15  
US-08-182-483A-17  
Sequence 17, Application US/08182483A  
Patent No. 5693486  
GENERAL INFORMATION:  
APPLICANT: LEHRER, ROBERT I.  
APPLICANT: KOKRYAKOV, VLADIMIR N.  
APPLICANT: HARTIG, SYLVIA S.L.  
TITLE OF INVENTION: NOVEL ANTIBIOTIC PEPTIDES  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1812  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/182,483A  
FILING DATE: 13-JAN-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 2000-0540.21  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030  
INFORMATION FOR SEQ ID NO: 17:



SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-182-483A-17

Query Match 62.0%; Score 31; DB 1; Length 18;  
Best Local Similarity 71.4%; Pred. No. 16;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGRKVC 8  
||| 1:1  
Db 2 GGRVC 8

Search completed: July 18, 1999, 06:07:57  
Job time: 299 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 06:07:05 ; Search time 44.47 Seconds

(without alignments)  
6.739 Million cell updates/sec

Title: US-09-142-043-7

Perfect score: 50

Sequence: 1 CGGKRVFC 8

Scoring table: BLOSUM62

Searched: 116738 segs, 37460341 residues

Database :

PIR\_58:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	37	74.0	527	2	A45611	probable hexose tr
2	37	74.0	140	2	E71439	hypothetical prote
3	35	70.0	248	1	PRMSC2	granzyme C (EC 3.4
4	35	70.0	1950	2	S12332	ubiquitin--protein
5	35	70.0	413	2	S46039	probable membrane
6	35	70.0	1700	2	S08167	Balbani ring 3 pr
7	34	68.0	26926	1	I38344	titin, cardiac mus
8	34	68.0	405	2	B30768	tryptophan synthas
9	34	68.0	408	2	S40715	hypothetical prote
10	33	66.0	63	2	S25772	testis-specific pr
11	33	66.0	880	2	S49627	regulatory protein
12	33	66.0	343	2	A69322	conserved hypothet
13	33	66.0	409	2	B71198	hypothetical prote
14	32	64.0	852	1	VCLJBR	env polypeptin -
15	32	64.0	464	2	A69417	cobyrilic acid synth
16	32	64.0	1021	2	E64576	hypothetical prote
17	32	64.0	295	2	S35293	hypothetical prote
18	32	64.0	537	2	B33485	spore coat protein
19	32	64.0	374	2	S69627	hypothetical prote
20	32	64.0	2014	2	I36936	complement recepto
21	32	64.0	661	2	I36937	complement recepto
22	32	64.0	2489	2	I73012	complement C3b/C4b
23	32	64.0	2039	2	A28507	complement C3b/C4b
24	32	64.0	1025	2	A54718	dihydropyrimidine
25	32	64.0	1025	2	B54718	dihydropyrimidine
26	31.5	63.0	312	2	A48413	probable olfactory
27	31.5	63.0	165	2	I38471	olfactory receptor
28	31.5	63.0	157	2	S58019	probable olfactory
29	31.5	63.0	157	2	S58033	probable olfactory
30	31	62.0	599	1	RKXBPV	RNA-directed RNA p
31	31	62.0	408	1	UQ2126	tryptophan synthas
32	31	62.0	401	1	TSPSBA	tryptophan synthas
33	31	62.0	90	1	R3RZ15	ribosomal protein
34	31	62.0	90	1	R3RZ15	ribosomal protein
35	31	62.0	431	2	S64704	ornithine decarbox
36	31	62.0	27	2	A58175	delta-conotoxin tx
37	31	62.0	90	2	A34435	ribosomal protein
38	31	62.0	521	2	A29345	steroid hormone re
39	31	62.0	292	2	A64086	regulatory protein

40	31	62.0	157	2	PN0685	hypothetical prote
41	31	62.0	404	2	S62475	hypothetical prote
42	31	62.0	1157	2	A55152	PAS1 protein - yea
43	31	62.0	1116	2	S57382	hypothetical prote
44	31	62.0	310	2	S31303	probable GTP-bind
45	31	62.0	1096	2	S62358	inositol 1,4,5-tri

## ALIGNMENTS

RESULT 1

A45611

Probable hexose transport protein - Trypanosoma brucei

C:Species: Trypanosoma brucei

C>Date: 22-Apr-1993 #sequence\_revision 18-Nov-1994 #text\_change 23-Mar-1995

C:Accession: A45611

R:Bringingaud, F.; Baltz, T.

Mol. Biochem. Parasitol. 52, 111-121, 1992

A>Title: A potential hexose transporter gene expressed predominantly in the bloodstre

A:Reference number: A45611; MUID:92326857

A:Accession: A45611

A>Status: Preliminary

A:Molecule type: mRNA

A:Residues: 1-527 <BRI>

A:Experimental source: subsp. brucei

A>Note: sequence inconsistent with the nucleotide translation

A>Note: sequence extracted from NCBI Backbone (NCBI:108238, NCBI:108239)

Query Match Best Local Similarity 74.0%; Score 37; DB 2; Length 527;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGKRVF 7  
Db 78 CGGKRVF 84

RESULT 2

E71439

hypothetical protein - Arabidopsis thaliana

C:Species: Arabidopsis thaliana

A:Variety: Columbia

C>Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 03-Aug-1998

C:Accession: E71439

R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D

P.; Medler, H.; Medler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terryn, N.; G

avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.

Nature 391, 485-488, 1998

A:Authors: Funk, B.; Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.;

H.; Duestenhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rec

Chauvier, C.; Chalmatis, N.

A>Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t

A:Reference number: A71400; MUID:98121113

A:Accession: E71439

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-140 <BEV>

A:Cross-references: GB:297342; NID:g2245031; PID:e327039; PID:g2245067

C:Genetics:

A:Map position: 4COP-4G3845

Query Match Best Local Similarity 74.0%; Score 37; DB 2; Length 140;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGKRVF 7  
Db 82 CGGKRVF 88

RESULT 3  
PRMSC2  
granzyme C (EC 3.4.21.-) precursor - mouse  
N:Alternate names: CPPI1; cytotoxic T-cell-specific proteinase 2; cytotoxic T-lymphocyte  
C:Species: Mus musculus (house mouse)  
C:Date: 30-Jun-1989 #sequence\_revision 14-Jul-1994 #text\_change 28-Aug-1998  
C:Accession: B28952; S03004; I49702; C26944  
R:Lobe, C.G.; Upton, C.; Duggan, B.; Ehrman, N.; Letellier, M.; Bell, J.; McPadden, G.;  
Biochemistry 27, 6941-6946, 1988  
A:Title: Organization of two genes encoding cytotoxic T lymphocyte-specific serine prote  
A:Reference number: A90536; M0ID:89062424  
A:Accession: B28952  
A:Molecule type: DNA  
A:Residues: 1-248 <LOB>  
A:Cross-references: GB:M22527; NID:g201028; PID:g201029  
R:Blanchley, R.C.; Duggan, B.; Ehrman, N.; Lobe, C.G.  
FEBS Lett. 234, 153-159, 1988  
A:Title: Isolation of two cDNA sequences which encode cytotoxic cell proteases.  
A:Reference number: S01006; M0ID:88271600  
A:Accession: S03004  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-248 <BLE>  
A:Cross-references: EMBL:X12822; NID:g50056; PID:g50057  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1988  
R:Jenne, D.E.; Rey, C.; Masson, D.; Stanley, K.K.; Hertz, J.; Plaetnick, G.; Tschoop, J.  
J. Immunol. 140, 318-323, 1988  
A:Title: cDNA cloning of granzyme C, a granule-associated serine protease of cytolytic T  
A:Reference number: I49702; M0ID:88088827  
A:Accession: I49702  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-68; 'R', 70-248 <JEN>  
A:Cross-references: GB:M18459; NID:g193621; PID:g309274  
R:Masson, D.; Tschoop, J.  
Cell 49, 679-685, 1987  
A:Title: A family of serine esterases in lytic granules of cytolytic T lymphocytes.  
A:Reference number: A90894; M0ID:87215932  
A:Accession: C26944  
A:Molecule type: protein  
A:Residues: 21-40 <MAS>  
C:Genetics:  
A:Introns: 19/1; 69/2; 114/3; 201/3  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; serine proteinase; T-cell  
F:1-18/Domain: signal sequence #status predicted <SIG>  
F:19-70/Domain: propeptide #status predicted <PPT>  
F:21-248/Product: granzyme C #status predicted <MPR>  
F:50-66,143-210,174-189/disulfide bonds: #status predicted  
F:65,109,204/Active site: His, Asp, Ser #status predicted

Query Match 70.0%; Score 35; DB 1; Length 248;  
Best Local Similarity 71.4%; Pred. No. 26;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGRKVC 8  
DB 44 GGRKVC 50

RESULT 4  
S12332  
ubiquitin--protein ligase (EC 6.3.2.19) - yeast (Saccharomyces cerevisiae) (strain S288C  
N:Alternate names: protein G7168; protein YGR184C  
C:Species: Saccharomyces cerevisiae  
C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 06-Feb-1998  
C:Accession: S12332; S64502; S64498  
R:Bartel, B.; Wuenning, I.; Varshavsky, A.  
EMBO J. 9, 3179-3189, 1990  
A:Title: The recognition component of the N-end rule pathway.  
A:Reference number: S12332; M0ID:91006011

A:Accession: S12332  
A:Molecule type: DNA  
A:Residues: 1-1950 <BAR>  
A:Cross-references: EMBL:X53747; NID:g4743; PID:g4744  
A:Experimental source: strain S288C  
R:Arroyo, J.; Garcia-Gonzalez, M.; Garcia-Saez, M.I.; Sanchez-Perez, M.; Nombela, C.  
submitted to the Protein Sequence Database, May 1996  
A:Reference number: S64499  
A:Accession: S64502  
A:Molecule type: DNA  
A:Residues: 1-1950 <ARC>  
A:Cross-references: EMBL:Z72969; NID:g1323325; PID:e243563; PID:g1323326; MIPS:YGR184  
A:Experimental source: strain S288C  
R:Hebling, U.; Hofmann, B.; Delius, H.  
submitted to the Protein Sequence Database, May 1996  
A:Reference number: S64403  
A:Accession: S64498  
A:Molecule type: DNA  
A:Residues: 1615-1950 <HEB>  
A:Cross-references: EMBL:Z72969; MIPS:YGR184C  
A:Experimental source: strain S288C  
C:Genetics:  
A:Gene: SGD:UBR1; PTR1  
A:Cross-references: SGD:S0003416; MIPS:YGR184C  
A:Map position: 7R  
C:Superfamily: ubiquitin--protein ligase  
C:Keywords: ligase; protein degradation

Query Match 70.0%; Score 35; DB 2; Length 1950;  
Best Local Similarity 85.7%; Pred. No. 14e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGSRKVC 7  
DB 1286 CGSRKVC 1292

RESULT 5  
S46039  
probable membrane protein YBR168w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein YBR1220  
C:Species: Saccharomyces cerevisiae  
C:Date: 26-Aug-1994 #sequence\_revision 09-Sep-1994 #text\_change 14-Nov-1997  
C:Accession: S46039  
R:Entian, K.D.; Koetter, P.; Rose, M.; Becker, J.; Grey, M.; Li, Z.; Niegemann, E.; S  
S.; Gruenbein, R.; Hedges, D.; Klesau, P.; Korol, S.; Krems, B.; Proft, M.; Siegers,  
submitted to the Protein Sequence Database, August 1994  
A:Reference number: S46013  
A:Accession: S46039  
A:Molecule type: DNA  
A:Residues: 1-413 <ENT>  
A:Cross-references: EMBL:Z36037; NID:g536509; PID:g536510; MIPS:YBR168w  
A:Experimental source: strain S288C  
C:Genetics:  
A:Map position: 2R  
C:Keywords: transmembrane protein  
F:36-52/Domain: transmembrane #status predicted <TM1>  
F:57-84/Domain: transmembrane #status predicted <TM2>  
F:104-128/Domain: transmembrane #status predicted <TM3>  
F:177-195/Domain: transmembrane #status predicted <TM4>

Query Match 70.0%; Score 35; DB 2; Length 413;  
Best Local Similarity 50.0%; Pred. No. 39;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGSRKVC 8  
DB 173 CGSRKVC 180

RESULT 6

S08167  
Baldiani ring 3 protein - midge (Chironomus tentans)  
C:Species: Chironomus tentans  
C>Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 20-Mar-1998  
C:Accession: S08167  
R:Paulsson, G.; Lendahl, U.; Galli, J.; Ericsson, C.; Wieslander, L.  
J. Mol. Biol. 211, 331-349, 1990  
A:Title: The baldiani ring 3 gene in Chironomus tentans has a diverged repetitive structure  
A:Reference number: S08167; MUID:90172404  
A:Accession: S08167  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-1700 <PAU>  
A:Cross-references: GB:X52263; NID:g7057; PID:g7058  
C:Genetics:  
A:Gene: BR3  
A:Map position: 4

Query Match 70.0%; Score 35; DB 2; Length 1700;  
Best Local Similarity 50.0%; Pred. No. 1.2e+02;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 CGGRKVC 8  
DB 1534 CGARKWC 1541

RESULT 7  
I38344  
titin, cardiac muscle - human  
N:Alternate names: connectin  
N:Contents: serine/threonine-specific protein kinase (EC 2.7.1.-)  
C:Species: Homo sapiens (man)  
C>Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 02-Jul-1998  
C:Accession: I38344; I38345; S20898; S20897; S6365; S37393  
R:Labell, S.; Kolmerer, B.  
Science 270, 293-296, 1995  
A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.  
A:Reference number: A57430; MUID:96026330  
A:Accession: I38344  
A:Status: nucleic acid sequence not shown; translated from GB/EMBL  
A:Molecule type: mRNA  
A:Residues: 1-26926 <LAB1>  
A:Cross-references: EMBL:X80568; NID:g1017424; PID:g1017425  
R:Musco, G.; Tziatzios, C.; Schmuck, P.; Pastore, A.  
Biochemistry 34, 553-561, 1995  
A:Title: Dissecting titin into its structural motifs: identification of an alpha-helix  
A:Reference number: I38345; MUID:95119041  
A:Accession: I38345  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1977-2014 <MUS>  
A:Cross-references: EMBL:X83270; NID:g602579; PID:g602580  
A:Note: conformation and properties are reported for a synthetic peptide corresponding to  
R:Labell, S.; Gautel, M.; Lakey, A.; Trinick, J.  
EMBO J. 11, 1711-1716, 1992  
A:Title: Towards a molecular understanding of titin.  
A:Reference number: S20897; MUID:92258380  
A:Accession: S20898  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 13597-14200, 'I', 14202-14696 <LAB2>  
A:Cross-references: EMBL:X64698; NID:g37192; PID:g37193  
A:Accession: S20897  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 16330-16382, 'S', 16384-16756, 'F', 16758-16860 <LAB3>  
A:Cross-references: EMBL:X64699; NID:g37190; PID:g37191  
A:Accession: S20899  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 'P', 22278-22431, 'R', 22433-22448, 'G', 22450-22453, 'Q', 22455-22480, 'TR', 22483-2

A:Cross-references: EMBL:X64697; NID:g37190; PID:g37195  
R:Kolmerer, B.; Olivieri, N.; Wilt, C.C.; Herrmann, B.G.; Labell, S.  
J. Mol. Biol. 256, 556-563, 1996  
A:Title: Genomic organization of M line titin and its tissue-specific expression in t  
A:Reference number: S63665; MUID:9617761  
A:Accession: S63665  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 26729-26825 <KOL>  
A:Cross-references: EMBL:X92412; NID:g1236761  
R:Gautel, M.; Leonard, K.; Labell, S.  
EMBO J. 12, 3827-3834, 1993  
A:Title: Phosphorylation of KSP motifs in the C-terminal region of titin in different  
A:Reference number: S37393; MUID:94008920  
A:Accession: S37393  
A:Molecule type: mRNA  
A:Residues: 26831-26926 <GAU>  
R:Improta, S.; Politou, A.S.; Pastore, A.  
submitted to the Brookhaven Protein Data Bank, February 1996  
A:Reference number: A66736; PDB:1FIT  
R:Pfuhl, M.; Pastore, A.  
submitted to the Brookhaven Protein Data Bank, August 1996  
A:Reference number: A66201; PDB:1NCT  
A:Contents: annotation; conformation by (1)H-NMR, residues 'S', 26059-26155  
C:Genetics:  
A:Gene: GDB:TTN  
A:Cross-references: GDB:127867; OMIM:188840  
A:Map position: 2q31-2q32  
C:Function:  
A:Description: structural protein forming filaments in striated muscle  
C:Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology;  
C:Keywords: alternative splicing; calmodulin binding; cardiac muscle; duplication; g1  
ral protein  
F:24753-25008/Domain: protein kinase homology <KIN>  
F:84, 177, 905, 2276, 2378, 2459, 2481, 2563, 2669, 2763, 2896, 3088, 3179, 3384, 3432, 3628, 3772, 40  
98, 11066, 11488, 11515, 11635, 11949, 12170, 12478, 12526, 12645, 12875, 13001, 13036, 13295, 1354  
F:16780, 16976, 17579, 17602, 17667, 17681, 17845, 17899, 18121, 18188, 18209, 18336, 18670, 18680  
F:1900, 21935, 22295, 22627, 22897, 23024, 2318, 23883, 24012, 24177, 24290, 24447, 24642,  
F:26171, 26178, 26184, 26190/Binding site: phosphate (Ser) (covalent) #status experiment

Query Match 68.0%; Score 34; DB 1; Length 26926;  
Best Local Similarity 83.3%; Pred. No. 1.7e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGGRKV 6  
DB 2687 CGGRKI 26872

RESULT 8  
B30768  
tryptophan synthase (EC 4.2.1.20) beta chain - Pseudomonas putida  
C:Species: Pseudomonas putida  
C>Date: 19-May-1989 #sequence\_revision 19-May-1989 #text\_change 31-Oct-1997  
C:Accession: S03835; B30768  
R:Crawford, I.P.  
submitted to the EMBL Data Library, October 1988  
A:Reference number: S03835  
A:Accession: S03835  
A:Molecule type: DNA  
A:Residues: 1-405 <CRA>  
A:Cross-references: EMBL:X13299; NID:g45748; PID:g45749  
C:Comment: This protein dimerizes to form the beta-2 subunit of tryptophan synthase.  
C:Genetics:  
A:Gene: trpB  
C:Superfamily: tryptophan synthase beta chain; tryptophan synthase beta chain homolog  
C:Keywords: carbon-oxygen lyase; hydro-lyase; pyridoxal phosphate  
F:12-395/Domain: tryptophan synthase beta chain homology <TRPB>  
F:95/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match 68.0%; Score 34; DB 2; Length 405;  
Best Local Similarity 71.4%; Pred. No. 57;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 CGRRKVC 7  
|||  
Db 76 CGGAKIF 82

## RESULT 9

hypothetical protein R10P11.3 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 09-Sep-1997  
C:Accession: S40715  
R:Almscough, R.; Mortimore, B.  
submitted to the EMBL Data Library, December 1993  
A:Reference number: S40713  
A:Accession: S40715  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-408 <ATN>  
A:Cross-references: EMBL:Z29095; NID:9436453; PID:9436456  
C:Genetics:  
A:Introns: 29/3; 64/1; 105/1; 252/3; 312/3

Query Match 68.0%; Score 34; DB 2; Length 408;  
Best Local Similarity 62.5%; Pred. No. 58;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 CGRRKVC 8  
|||  
Db 249 CGDQKVC 256

## RESULT 10

S25772  
testis-specific protein Mst84da - fruit fly (*Drosophila melanogaster*)  
C:Species: *Drosophila melanogaster*  
C:Date: 26-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 24-Sep-1998  
C:Accession: S25772; A56565  
R:Kuhn, R.; Kuhn, C.; Boersch, D.; Glaetzer, K.H.; Schaefer, U.; Schaefer, M.  
Mech. Dev. 35, 143-151, 1991  
A:Title: A cluster of four genes selectively expressed in the male germ line of *Drosophila*  
A:Reference number: A56565; MUID:92102953  
A:Accession: S25772  
A:Molecule type: DNA  
A:Residues: 1-63 <KUH>  
A:Cross-references: EMBL:X67703; NID:q11072; PID:q11073  
A:Note: sequence extracted from NCBI backbone (NCBI:74219, NCBI:74224)  
C:Genetics:  
A:Gene: Mst84da  
A:Cross-references: FlyBase:FBgn0004172  
A:Map position: 3  
A:Introns: 13/3  
C:Superfamily: fruit fly testis-specific protein  
C:Keywords: spermatogenesis; tandem repeat

Query Match 66.0%; Score 33; DB 2; Length 63;  
Best Local Similarity 62.5%; Pred. No. 19;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 CGRRKVC 8  
|||  
Db 50 CGTSSFC 57

RESULT 11  
S4967  
regulatory protein ARGR2 - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: protein YML099c  
C:Species: *Saccharomyces cerevisiae*  
C:Date: 02-Dec-1994 #sequence\_revision 10-Feb-1995 #text\_change 12-Dec-1997  
C:Accession: S49627; A25064  
R:Gentiles, S.; Borman, S.  
submitted to the EMBL Data Library, November 1994  
A:Reference number: S49627  
A:Accession: S49627  
A:Molecule type: DNA  
A:Residues: 1-880 <GEN>

A:Cross-references: EMBL:Z46660; NID:9575702; PID:9575703; MIPS:YML099c  
R:Messenguy, F.; Dubois, E.; Descamps, F.  
Eur. J. Biochem. 157, 77-81, 1986  
A:Title: Nucleotide sequence of the ARGR1 regulatory gene and amino acid sequence ho  
A:Reference number: A25064; MUID:86220196  
A:Accession: A25064

A:Molecule type: DNA  
A:Residues: 1-3, 'F', 'S', '128-131, 'L', '132-282, 'V', '284-344, 'V', '346-365, 'Q', '367-548, 'A'  
C:Genetics:  
A:Gene: SGD:ARG81; ARGR2  
A:Cross-references: MIPS:YML099c; SGD:S0004565  
A:Map position: 13L  
C:Superfamily: unassigned GAL4-type zinc cluster proteins: GAL4 zinc binuclear cluste  
C:Keywords: DNA binding; nucleus  
F:16-53/Domain: GAL4 zinc binuclear cluster homology <GAL4>

Query Match 66.0%; Score 33; DB 2; Length 880;  
Best Local Similarity 75.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 CGRRKVC 8  
|||  
Db 24 CGRRKVC 31

## RESULT 12

A69322  
conserved hypothetical protein AF0577 - *Archaeoglobus fulgidus*  
C:Species: *Archaeoglobus fulgidus*  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 05-Jun-1998  
C:Accession: A69322  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
A:; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.C.; Gill, S.; Kirkness, E  
Nature 390, 364-370, 1997  
A:Authors: Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P  
ese, C.R.; Venter, J.C.  
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch  
A:Reference number: A69250; MUID:98049343  
A:Accession: A69322  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-343 <KLE>  
A:Cross-references: GB:AE001065; GB:AE000782; NID:92689388; PID:92650048; TIGR:AF0577

Query Match 66.0%; Score 33; DB 2; Length 343;  
Best Local Similarity 85.7%; Pred. No. 76;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CGRRKVC 7  
|||  
Db 284 CGRRKVC 290

## RESULT 13

B71198  
hypothetical protein PH1856 - *Pyrococcus horikoshii*  
C:Species: *Pyrococcus horikoshii*  
C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 14-Aug-1998  
C:Accession: B71198  
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Se

M.: Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushiida, N.; Oguchi  
DNA Res. 5, 55-76, 1998  
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic  
A:Reference number: A71000  
A:Accession: B71198  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-409 <RAW>  
A:Cross-references: GB:AP000007; NID:g3236134; PID:d1031920; PID:g3258294  
A:Experimental source: strain OT3  
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank  
C:Genetics:  
A:gene: PH1856

Query Match 66.0%; Score 33; DB 2; Length 409;  
Best Local Similarity 62.5%; Pred. No. 87;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 CGGRKVC 8  
| | | | |  
Db 12 CKGRKLC 19

## RESULT 14

env polyprotein - human immunodeficiency virus type 1 (isolate BR)  
N:Alternate names: coat polyprotein  
N:Contains: coat protein gp120; coat protein gp41  
C:Species: human immunodeficiency virus type 1, HIV-1  
A:Note: host Homo sapiens (man)  
C:Date: 31-Dec-1989 #sequence\_revision 31-Dec-1989 #text\_change 16-Feb-1997  
C:Accession: A31667  
R:Rand, R.; Thayer, R.; Srinivasan, A.; Nayyar, S.; Gardner, M.; Luciw, P.; Dandekar, S.  
Virology 168, 79-89, 1989  
A:Title: Biological and molecular characterization of human immunodeficiency virus (HIV-  
A:Reference number: A94389; M0ID:89085613  
A:Accession: A31667  
A:Molecule type: DNA  
A:Residues: 1-852 <ANA>  
C:Superfamily: type E retrovirus env polyprotein  
C:Keywords: capsid protein; coat protein; polyprotein; transmembrane protein  
F:1-516/Product: coat protein gp120 #status predicted <CP2>  
F:517-852/Product: coat protein gp41 #status predicted <CP2>

Query Match 64.0%; Score 32; DB 1; Length 852;  
Best Local Similarity 62.5%; Pred. No. 2.4e+02;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 CGGRKVC 8  
| | | | |  
Db 376 CGGEFFC 383

## RESULT 15

A69417  
cobyrilic acid synthase (cbp) homolog - Archaeoglobus fulgidus  
C:Species: Archaeoglobus fulgidus  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 18-Sep-1998  
C:Accession: A69417  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;  
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.; Utterbe  
Nature 390, 364-370, 1997  
A:Authors: Colton, M.D.; Springs, T.; Artlich, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.  
ese, C.R.; Venter, J.C.  
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo  
A:Reference number: A69250; M0ID:98049343  
A:Accession: A69417  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-464 <KLE>

A:Cross-references: GB:AE001011; GB:AE000782; NID:g2689334; PID:g2649236; TIGR:AF1338  
C:Superfamily: probable cobyrilic acid synthase

Query Match 64.0%; Score 32; DB 2; Length 464;  
Best Local Similarity 50.0%; Pred. No. 1.5e+02;  
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 CGGRKVC 8  
| | | | |  
Db 321 CGGRKVC 328

Search completed: July 18, 1999, 06:07:06  
Job time: 252 sec





GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 11:37:52 ; Search time 31.53 Seconds  
(without alignments)  
6.810 Million cell updates/sec

Title: US-09-142-043-7

Perfect score: 50

Sequence: 1 CGGRKVF 8

Scoring table: BLOSUM62

Searched: 74019 seqs, 26840295 residues

Database : SwissProt\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	37	74.0	527	1	TH11_TRYBB	Q06221 trypanosoma
2	37	74.0	528	1	TH12_TRYBB	Q09037 trypanosoma
3	35	70.0	1700	1	BAR3_CHITE	Q03376 chironomus
4	35	70.0	248	1	GRAC_MOUSE	P08882 mus musculus
5	35	70.0	1950	1	UBR1_YEAST	P18912 saccharomyc
6	35	70.0	413	1	YB18_YEAST	P38922 saccharomyc
7	34	68.0	405	1	TRPB_PSEPU	P11080 pseudomonas
8	34	68.0	408	1	UBPX_CAEEL	P34547 caenorhabdi
9	33	66.0	880	1	ARG2_YEAST	P05085 saccharomyc
10	33	66.0	63	1	M84A_DROME	Q01642 drosophila
11	32	64.0	2039	1	CRL1_HUMAN	P17827 homo sapien
12	32	64.0	1025	1	DPYD_BOVIN	Q28007 bos taurus
13	32	64.0	1025	1	DPYD_HUMAN	Q12882 homo sapien
14	32	64.0	1025	1	DPYD_PIG	Q28943 sus scrofa
15	32	64.0	852	1	ENV_HVIBN	P12488 human immun
16	32	64.0	122	1	IPSG_MEIME	P16226 melles melles
17	32	64.0	537	1	SP70_DICDI	P15629 dictyosteli
18	32	64.0	197	1	VE26_HNVOP	P15629 dictyosteli
19	31.5	63.0	312	1	OLFE_HUMAN	O10276 oryza pseu
20	31	62.0	431	1	DCOR_DATST	P34982 homo sapien
21	31	62.0	521	1	ERR1_HUMAN	P50134 datura stra
22	31	62.0	310	1	CTRI_YEAST	P11474 homo sapien
23	31	62.0	292	1	ILVY_YEAST	Q00882 saccharomyc
24	31	62.0	85	1	IT15_HAEIN	P44621 haemophilus
25	31	62.0	1157	1	IT15_ARATH	O22867 arabidopsis
26	31	62.0	203	1	RL15_ANOGA	P52818 anopheles g
27	31	62.0	90	1	RL15_MAIZE	P17703 zea mays (m
28	31	62.0	90	1	RL15_ORYSA	P12550 oryza sativ
29	31	62.0	90	1	RL15_SECEA	P20283 secale cere
30	31	62.0	599	1	RPO_PEAUV	P29154 pea enation
31	31	62.0	402	1	TRPB_PSEAE	P07345 pseudomonas
32	31	62.0	408	1	TRPB_PSESY	P34617 pseudomonas
33	31	62.0	404	1	YACS_SCHPO	Q09819 schizosacch
34	31	62.0	571	1	YB63_SCHPO	Q09744 schizosacch
35	30.5	61.0	125	1	LECC_POLMI	P16108 polyanthroc
36	30	60.0	1656	1	ATC8_YEAST	Q12674 saccharomyc
37	30	60.0	68	1	CSPA_STIAU	P72366 stigmatella
38	30	60.0	160	1	DYR_MYGE	P47470 mycoplasma
39	30	60.0	200	1	EIA_ADEM1	P12534 mouse adeno
40	30	60.0	457	1	FKB5_HUMAN	O13551 homo sapien
41	30	60.0	456	1	FKB5_MOUSE	O64578 mus musculu
42	30	60.0	161	1	GSPN_ERWCH	Q47422 erwilia chr
43	30	60.0	84	1	MT22_ORYSA	P93433 oryza sativ

## ALIGNMENTS

RESULT	1	TH11_TRYBB	STANDARD	PRT	527 AA.	Q55200 synecocyst
AC	006221	TH11_TRYBB				P23491 mesocricetu
DT	01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)					
DT	01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)					
DT	01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)					
DE	GLUCOSE TRANSPORTER 1B/1C/1D/1E/2B.					
GN	TH11B AND TH11C AND TH11D AND TH11F AND TH11G.					
OS	TRYPANOSOMA BRUCEI BRUCEI.					
OC	EUKARYOTA; PROTOZOA; SARCOMASTIGOPHORA; MASTIGOPHORA; KINETOPLASTIDA;					
OC	TRYPANOSOMATIDAE.					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=ATRO 164.					
RX	MEDLINE: 93140752.					
RA	BRINGAUD F., BALIZ T.;					
RL	MOL. CELL. BIOL. 13:1146-1154(1993).					
CC	- FUNCTION: FACILITATIVE GLUCOSE TRANSPORTER.					
CC	- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.					
CC	- DEVELOPMENTAL STAGE: EXPRESSED SPECIFICALLY IN BLOODSTREAM FORMS.					
CC	- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.					
DR	EMBL: X69091; E56572; -					
DR	EMBL: X69091; E56573; -					
DR	EMBL: X69091; E56574; -					
DR	EMBL: X69091; E56576; -					
DR	PROSITE: P800216; SUGAR_TRANSPORT_1; FALSE_NEG.					
DR	PROSITE: P800217; SUGAR_TRANSPORT_2; FALSE_NEG.					
KW	SUGAR_TRANSPORT; TRANSMEMBRANE; MULTIGENE FAMILY.					
FT	TRANSMEM	37	62			
FT	DOMAIN	63	111			
FT	TRANSMEM	112	137			
FT	TRANSMEM	151	170			
FT	TRANSMEM	181	200			
FT	TRANSMEM	212	234			
FT	TRANSMEM	248	268			
FT	DOMAIN	269	293			
FT	TRANSMEM	294	314			
FT	TRANSMEM	338	359			
FT	TRANSMEM	364	391			
FT	TRANSMEM	399	422			
FT	TRANSMEM	437	458			
FT	TRANSMEM	473	492			
FT	DOMAIN	493	527			
SQ	SEQUENCE	527 AA;	56469 MW;	AAD152EA CRC32;		
QY	1 CGGRKVF 7					
DB	78 CGGRKVF 84					
RESULT	2	TH12_TRYBB	STANDARD	PRT	528 AA.	
AC	009037	TH12_TRYBB				
DT	01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)					
DT	01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)					
DT	01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)					
DE	GLUCOSE TRANSPORTER 1E.					
GN	TH11E.					

OS TRYPAOSOMA BRUCEI BRUCEI  
 OC EUKARYOTA: PROTOSTOMA: SARCOMASTIGOPHORA: MASTIGOPHORA: KINETOPLASTIDA;  
 CC TRYPAOSOMATIDAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-EATRO 164;  
 RX MEDLINE: 93140752.  
 RA BRINGAUD F., BALTZ T.;  
 RL MOL. CELL. BIOL. 13:1146-1154(1993).  
 CC -1- FUNCTION: FACILITATIVE GLUCOSE TRANSPORTER.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED SPECIFICALLY IN BLOODSTREAM FORMS.  
 CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
 DR EMBL: X69091; G312954; -;  
 DR PROSITE: PS00216; SUGAR\_TRANSPORT\_1; FALSE\_NEG.  
 DR PROSITE: PS00217; SUGAR\_TRANSPORT\_2; FALSE\_NEG.  
 KM SUGAR TRANSPORT; TRANSMEMBRANE; MULTIGENE FAMILY.  
 FT TRANSSEM 37 62 POTENTIAL.  
 FT DOMAIN 63 111 EXTRACELLULAR (POTENTIAL).  
 FT TRANSSEM 112 137 POTENTIAL.  
 FT TRANSSEM 151 170 POTENTIAL.  
 FT TRANSSEM 181 200 POTENTIAL.  
 FT TRANSSEM 212 234 POTENTIAL.  
 FT TRANSSEM 248 268 POTENTIAL.  
 FT DOMAIN 269 293 CYTOPLASMIC (POTENTIAL).  
 FT TRANSSEM 294 314 POTENTIAL.  
 FT TRANSSEM 338 359 POTENTIAL.  
 FT TRANSSEM 364 391 POTENTIAL.  
 FT TRANSSEM 399 422 POTENTIAL.  
 FT TRANSSEM 437 458 POTENTIAL.  
 FT TRANSSEM 473 492 POTENTIAL.  
 FT DOMAIN 493 528 CYTOPLASMIC (POTENTIAL).  
 SQ SEQUENCE 528 AA; 56827 MW; 5BCBFD8C CRC32;

Query Match 74.08; Score 37; DB 1; Length 528;  
 Best Local Similarity 71.46; Pred. No. 13;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGRKVF 7  
 Db 78 CGGRKIV 84

RESULT 3  
 BAR3\_CHITE STANDARD; PRT: 1700 AA.  
 ID BAR3\_CHITE  
 AC 003376;  
 DT 01-OCT-1993 (REL. 27, CREATED)  
 DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)  
 DE BALBIANI RING PROTEIN 3 PRECURSOR.  
 GN BR3.  
 OS CHIRONOMUS TENTANS (MIDGE).  
 OC EUKARYOTA: METAZOA: ARTHROPODA: INSECTA: DIPTERA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-SALIVARY GLAND;  
 RX MEDLINE: 90172404.  
 RA PAULSSON G., LENDAHU U., GALLI J., ERICSSON C., WIELANDER L.;  
 RL J. MOL. BIOL. 211:331-349(1990).  
 CC -1- FUNCTION: USED BY THE LARVAE TO CONSTRUCT A SUPRAMOLECULAR  
 CC STRUCTURE, THE LARVAL TUBE. BALBIANI RING PROTEIN 3 COULD PLAY A  
 CC ROLE AS A TRANSPORT PROTEIN THAT BINDS TO OTHER PROTEINS  
 CC INTRACELLULARLY AND IN THE GLAND LUMEN IN ORDER TO PREVENT THESE  
 CC FROM FORMING WATER-INSOLUBLE FIBERS TOO EARLY.  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- TISSUE SPECIFICITY: SALIVARY GLAND.  
 CC -1- DOMAIN: HAS 82 APPROXIMATE REPEATS OF CYS-X-CYS-X-CYS.  
 DR EMBL: X52263; G7058; -;  
 DR PIR: S08167; S08167.  
 DR HSSP: P01215; 1HCN.  
 KM REPEAT; SIGNAL.

FT SIGNAL 1 2 POTENTIAL.  
 FT CHAIN ? 1700 BALBIANI RING PROTEIN 3.  
 SQ SEQUENCE 1700 AA; 186145 MW; D5A2BC28 CRC32;

Query Match 70.08; Score 35; DB 1; Length 1700;  
 Best Local Similarity 50.0%; Pred. No. 79;  
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGGRKVF 8  
 Db 1534 CGAKKIVC 1541

RESULT 4  
 GRAC\_MOUSE STANDARD; PRT: 248 AA.  
 ID GRAC\_MOUSE  
 AC P08882;  
 DT 01-NOV-1988 (REL. 09, CREATED)  
 DT 01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE GRANZYME C PRECURSOR (EC 3.4.21.-) (CYTOTOXIC CELL PROTEASE 2) (CCP2)  
 DE (B10).  
 GN GZMC OR CTLA5 OR CTLA-5.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA:  
 CC EUTHERIA: RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 88271600.  
 RA BLACKLEY R.C., DUGGAN B., EHMAN N., LOBE C.G.;  
 RL FEBS LETT. 234:153-159(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 89062424.  
 RA LOBE C.G., UFTON C., DUGGAN B., EHMAN N., LETELLIER M., BELL J.,  
 RA MCFADEN G., BLACKLEY R.C.;  
 RL BIOCHEMISTRY 27:6941-6946(1988).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 88088827.  
 RA JENNE D.E., REY C., MASSON D., STANLEY K.K., HERZ J., PLATTINCK G.,  
 RA TSCHOPP J.;  
 RL J. IMMUNOL. 140:318-323(1988).  
 RN [4]  
 RP SEQUENCE OF 21-40.  
 RX MEDLINE: 87215932.  
 RA MASSON D., TSCHOPP J.;  
 RL CELL 49:679-685(1987).  
 CC -1- FUNCTION: THIS ENZYME IS PROBABLY NECESSARY FOR TARGET CELL  
 CC LYSIS IN CELL-MEDIATED IMMUNE RESPONSES.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC GRANULES OF CYTOLYTIC  
 CC T-LYMPHOCYTES.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY. STRONGEST TO OTHER GRANZYMES AND TO MAST CELL  
 CC PROTEASES.  
 DR EMBL: M22527; G201029; -;  
 DR EMBL: X12823; G50057; -;  
 DR EMBL: M18459; G309274; -;  
 DR PIR: B28952; B28952.  
 DR PIR: C26944; C26944.  
 DR PIR: S03004; S03004.  
 DR HSSP: P00761; 1EPT.  
 DR MGD: MGT109256; GZMC.  
 DR PROSITE: PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE: PS00135; TRYPSIN\_SER; FALSE\_NEG.  
 KM HYDROLASE; SERINE PROTEASE; ZYMOGEN; SIGNAL; T-CELL; CYTOLYSIS.  
 FT SIGNAL 1 18  
 FT PROPEP 19 20  
 FT CHAIN 21 248  
 FT ACT\_SITE 65 65 GRANZYME C.  
 FT ACT\_SITE 109 109 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 204 204 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT DISULFID 50 66 BY SIMILARITY.  
FT DISULFID 143 210 BY SIMILARITY.  
FT DISULFID 174 189 BY SIMILARITY.  
FT CONFLICT 69 69 S -> R (IN REF. 3).  
SQ SEQUENCE 248 AA; 27310 MM; 63CE6727 CRC32;

Query Match 70.0%; Score 35; DB 1; Length 248;  
Best Local Similarity 71.4%; Pred. No. 16;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 CGRKYFC 8  
DB 44 CGRKYFC 50

RESULT 5  
UBRL\_YEAST  
ID UBRL\_YEAST STANDARD; PRT; 1950 AA.  
AC P19812;  
DT 01-FEB-1991 (REL. 17, CREATED)  
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE N-END-RECOGNIZING PROTEIN (UBIQUITIN-PROTEIN LIGASE E3 COMPONENT) (N-  
DE RECOGNITN).  
GN UBRL OR PTR1 OR YGR184C OR G7168.  
OS SACHAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C / GRF88;  
RX MEDLINE: 91006011.  
RA BARTEL B., WUENNING I., VARSHAVSKY A.;  
RL EMBO J. 9:3179-3189(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C;  
RX MEDLINE: 97279231.  
RA ARROYO J., GARCIA-GONZALEZ M., GARCIA-SAEZ M.I., SANCHEZ-PEREZ M.,  
RA NOMBELA C.;  
RL YEAST 13:357-363(1997).  
RN [3]  
RP SEQUENCE OF 346-1950 FROM N.A.  
RA HEBLING U., HOFMANN B., DELIUS H.;  
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DDJ DATA BANKS.  
CC -1- FUNCTION: UBRL IS THE RECOGNITION COMPONENT OF THE N-END RULE  
PATHWAY. UBRL BINDS TO PROTEINS BEARING AMINO-TERMINAL RESIDUES  
THAT ARE DESTABILIZING ACCORDING TO THE N-END RULE, BUT DOES NOT  
CC BIND TO OTHERWISE IDENTICAL PROTEINS BEARING STABILIZING AMINO-  
CC TERMINAL RESIDUES.  
CC EMBL: X53747; G4744; -;  
DR EMBL: X99074; E252628; -;  
DR EMBL: 272969; E245563; -;  
DR PIR: S12332; S12332.  
DR SGD: L0002420; UBRL.  
KM UBIQUITIN CONJUGATION.  
SQ SEQUENCE 1950 AA; 224836 MM; FC3BA78D CRC32;

Query Match 70.0%; Score 35; DB 1; Length 1950;  
Best Local Similarity 85.7%; Pred. No. 89;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CGRKYFC 7  
DB 1286 CGRKYFC 1292

RESULT 6  
YB18\_YEAST  
ID YB18\_YEAST STANDARD; PRT; 413 AA.  
AC P38392;  
DT 01-OCT-1994 (REL. 30, CREATED)

DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)  
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 48.6 KD PROTEIN IN TYR1-SSE2 INTERGENIC REGION.  
GN YB186W OR YB1220.  
OS SACHAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C;  
RA ENTIAN K.-D., KOETTER P., ROSE M., BECKER J., GREY M., LI Z.,  
RA NIEGEMANN E., SCHENK-GROENINGER R., SEROS J., WENNER E.,  
RA WOLTER R., BRENDL M., BAUER J., BRAUN H., DERN K., DUESTERHUS S.,  
RA GRUENBEIN R., HEDGES D., KIESAU P., KOROL S., KREWS B., PROFIT M.,  
RA SIEGERS K., BAUR A., BOLES E., MIOSGA T.,  
RA SCHAFF-GERSTENSCHLAGER I., ZIMMERMANN F.K.;  
RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DDJ DATA BANKS.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
DR EMBL: Z36037; G536510; -;  
DR PIR: S46039; S46039.  
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.

FT TRANSMEM 67 87 POTENTIAL.  
FT TRANSMEM 104 124 POTENTIAL.  
FT TRANSMEM 182 202 POTENTIAL.  
SQ SEQUENCE 413 AA; 48578 MM; 9F1266C2 CRC32;

Query Match 70.0%; Score 35; DB 1; Length 413;  
Best Local Similarity 50.0%; Pred. No. 24;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 CGRKYFC 8  
DB 173 CGRKYFC 180

RESULT 7  
TRPB\_PSEPU  
ID TRPB\_PSEPU STANDARD; PRT; 405 AA.  
AC P11080;  
DT 01-JUL-1989 (REL. 11, CREATED)  
DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)  
DT 01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)  
DE TRYPTOPHAN SYNTHASE BETA CHAIN (EC 4.2.1.20).  
GN TRPB.  
OS PSEUDOMONAS PUTIDA.  
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;  
OC PSEUDOMONADACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PG1 C1S;  
RX MEDLINE: 89335826.  
RA EBERLY L., CRAWFORD I.P.;  
RL BIOCHIMIE 71:521-531(1989).  
CC -1- FUNCTION: THE BETA SUBUNIT IS RESPONSIBLE FOR THE THE SYNTHESIS  
OF L-TRYPTOPHAN FROM INDOLE AND L-SERINE.  
CC -1- CATALYTIC ACTIVITY: L-SERINE + 1-(INDOL-3-YL)GLYCEROL 3-PHOSPHATE  
CC = L-TRYPTOPHAN + GLYCERALDEHYDE 3-PHOSPHATE + H(2)O.  
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.  
CC -1- PATHWAY: LAST (FIFTH) STEP IN BIOSYNTHESIS OF TRYPTOPHAN.  
CC -1- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS.  
DR EMBL: X13299; G45749; -;  
DR PIR: B30768; B30768.  
DR PIR: S03835; S03835.  
DR HSR: P00933; IMST.  
DR PROSITE: PS00168; TRP-SYNTHASE.BETA.1.  
KM TRYPTOPHAN BIOSYNTHESIS; LYASE.  
FT BINDING 95 95 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
SQ SEQUENCE 405 AA; 44047 MM; B8CA769D CRC32;

Query Match 68.0%; Score 34; DB 1; Length 405;  
Best Local Similarity 71.4%; Pred. No. 36;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 CGGRKVF 7  
111111  
Db 76 CGGAKIF 82

RESULT 8  
UBPX\_CAEEL STANDARD; PRT; 408 AA.

AC P34547;  
DT 01-FEB-1994 (REL. 28, CREATED)  
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE PROBABLE UBIQUITIN CARBOXYL-TERMINAL HYDROLASE R10E11.3 (EC 3.1.2.15)  
DE (UBIQUITIN THIOLESTERASE) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE)  
GN R10E11.3  
OS CAENORHABDITIS ELEGANS.

OC EUKARYOTA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE: 94150718.

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COSEY T., COOPER J., COULSON A.,  
RA CRAWTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,  
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,  
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSER N.,  
RA LARREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SANDERS D., SHOWNKEEN R.,  
RA SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,  
RA SULTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,  
RA WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,  
RA WOLDMAN P.;  
RL NATURE 368:32-38(1994).  
CC -1- CATALYTIC ACTIVITY: UBIQUITIN C-TERMINAL THIOLESTER + H(2)O =  
CC -1- UBIQUITIN + A THIOL.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C19; ALSO KNOWN AS  
CC FAMILY 2 OF UBIQUITIN CARBOXYL-TERMINAL HYDROLASES.  
DR EMBL: 229095; GA36456; -.  
DR PIR: S40715; S40715.  
DR WORMPEP; R10E11.3; CE00305.  
DR PROSITE; PS00972; UCH\_2\_1; 1.  
DR PROSITE; PS00973; UCH\_2\_2; 1.  
KW HYPOTHETICAL PROTEIN; UBIQUITIN CONJUGATION; HYDROLASE;  
KW THIOL PROTEASE.  
FT ACT\_SITE 34 BY SIMILARITY.  
FT ACT\_SITE 331 BY SIMILARITY.  
FT ACT\_SITE 340 BY SIMILARITY.  
SQ SEQUENCE 408 AA; 46356 MW; 6C6E014D CRC32;

Query Match 68.0%; Score 34; DB 1; Length 408;  
Best Local Similarity 62.5%; Pred. NO. 36;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 CGGRKVF 8  
111111  
Db 249 CGDQKIF 256

RESULT 9  
ARG2\_YEAST STANDARD; PRT; 880 AA.

AC P05085;  
DT 01-JAN-1988 (REL. 06, CREATED)  
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE ARGININE METABOLISM REGULATION PROTEIN II.  
GN ARG2 OR ARG81 OR YML099C.  
OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.  
RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE; 86220196.  
RA MESSENGUY F., DUBOIS E., DESCAMPS F.;  
RL EUR. J. BIOCHEM. 157:77-81(1986).  
RN [2]

RP SEQUENCE FROM N.A.  
RC STRAIN-S288C / AB972;  
RA GENTLES S., BOWMAN S., BARRELL B.G., RAJANDREAM M.A.;  
RL SUBMITTED (NOV-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- FUNCTION: THIS PROTEIN REPRESS OR INDICES SEVERAL GENES OR  
CC ARGININE METABOLISM.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- SIMILARITY: CONTAINS A ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR  
CC CLUSTER DOMAIN.

DR EMBL: X03940; G3378; -.  
DR EMBL: 246560; G575703; -.  
DR PIR: A25064; A25064.  
DR TRNSPAC; T00044; -.

DR SGD; L0000114; ARG82.

DR PROSITE; PS00463; ZN2\_CYS6\_FUNGAL\_1; 1.  
DR PROSITE; PS50048; ZN2\_CYS6\_FUNGAL\_2; 1.

KW TRANSCRIPTION REGULATION; ACTIVATOR; DNA-BINDING; NUCLEAR PROTEIN;  
KW ZINC; METAL-BINDING; ARGININE METABOLISM.

FT DNA\_BIND 21 48 ZN(2)-CYS(6), FUNGAL-TYPE.  
FT SIMILAR 587 690 TO N.C.RASSA NMR (AA 96-193).  
FT CONFLICT 4 4 S -> F (IN REF. 1).  
FT CONFLICT 129 129 MISSING (IN REF. 1).  
FT CONFLICT 283 283 F -> V (IN REF. 1).  
FT CONFLICT 345 345 D -> V (IN REF. 1).  
FT CONFLICT 366 366 E -> Q (IN REF. 1).  
FT CONFLICT 549 549 T -> A (IN REF. 1).  
FT CONFLICT 597 597 K -> S (IN REF. 1).  
FT CONFLICT 665 665 K -> N (IN REF. 1).  
FT CONFLICT 869 869 V -> I (IN REF. 1).  
SQ SEQUENCE 880 AA; 100281 MW; CDBA1C8D CRC32;

Query Match 66.0%; Score 33; DB 1; Length 880;  
Best Local Similarity 75.0%; Pred. NO. 1e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 CGGRKVF 8  
111111  
Db 24 CGGRKVF 31

RESULT 10  
M84A\_DROME STANDARD; PRT; 63 AA.

AC 001642;  
DT 01-JUL-1993 (REL. 26, CREATED)  
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE MALE SPECIFIC SPERM PROTEIN MST84DA.  
GN MST84DA.

OS DROSOPHILA MELANOGASTER (FRUIT FLY).  
OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-OREGON-R;  
RX MEDLINE: 92102953.  
RA KUHN R., KUHN C., BOERSCH D., GLAETZER K.H., SCHAEFER U.,  
RA SCHAEFER M.;  
RL MECH. DEV. 35:143-151(1991).

CC -1- TISSUE SPECIFICITY: TESTIS.  
CC -1- DEVELOPMENTAL STAGE: PRIMARY SPERMATOCYTES.  
CC -1- DOMAIN: THIS PROTEIN IS MOSTLY COMPOSED OF REPETITIVE C-G-P  
CC MOTIFS.

CC -1- SIMILARITY: BELONGS TO THE MST(3)CGP FAMILY.  
DR EMBL: X67703; G11073; -.  
DR EMBL: F890004172; M8484DA.  
DR EMBL: F890004172; M8484DA.  
KW SPERMATOCYTES; REPEAT; MULTIGENE FAMILY.

SQ SEQUENCE 63 AA; 5806 MW; 4CAF7891 CRC32;



```

FT DISULFID 1095 1132 BY SIMILARITY.
FT DISULFID 1138 1180 BY SIMILARITY.
FT DISULFID 1166 1193 BY SIMILARITY.
FT DISULFID 1197 1240 BY SIMILARITY.
FT DISULFID 1226 1253 BY SIMILARITY.
FT DISULFID 1258 1300 BY SIMILARITY.
FT DISULFID 1286 1316 BY SIMILARITY.
FT DISULFID 1321 1370 BY SIMILARITY.
FT DISULFID 1350 1387 BY SIMILARITY.
FT DISULFID 1396 1439 BY SIMILARITY.
FT DISULFID 1426 1452 BY SIMILARITY.
FT DISULFID 1457 1498 BY SIMILARITY.
FT DISULFID 1484 1514 BY SIMILARITY.
FT DISULFID 1519 1568 BY SIMILARITY.
FT DISULFID 1548 1585 BY SIMILARITY.
FT DISULFID 1591 1633 BY SIMILARITY.
FT DISULFID 1619 1646 BY SIMILARITY.
FT DISULFID 1650 1693 BY SIMILARITY.
FT DISULFID 1679 1706 BY SIMILARITY.
FT DISULFID 1711 1753 BY SIMILARITY.
FT DISULFID 1739 1769 BY SIMILARITY.
FT DISULFID 1774 1823 BY SIMILARITY.
FT DISULFID 1803 1840 BY SIMILARITY.
FT DISULFID 1848 1891 BY SIMILARITY.
FT DISULFID 1877 1904 BY SIMILARITY.
FT DISULFID 1909 1952 BY SIMILARITY.
FT DISULFID 1938 1965 BY SIMILARITY.
FT CARBOHYD 56 56 POTENTIAL.
FT CARBOHYD 252 252 POTENTIAL.
FT CARBOHYD 410 410 POTENTIAL.
FT CARBOHYD 447 447 POTENTIAL.
FT CARBOHYD 509 509 POTENTIAL.
FT CARBOHYD 578 578 POTENTIAL.
FT CARBOHYD 702 702 POTENTIAL.
FT CARBOHYD 860 860 POTENTIAL.
FT CARBOHYD 897 897 POTENTIAL.
FT CARBOHYD 959 959 POTENTIAL.
FT CARBOHYD 1028 1028 POTENTIAL.
FT CARBOHYD 1152 1152 POTENTIAL.
FT CARBOHYD 1310 1310 POTENTIAL.
FT CARBOHYD 1481 1481 POTENTIAL.
FT CARBOHYD 1504 1504 POTENTIAL.
FT CARBOHYD 1534 1534 POTENTIAL.
FT CARBOHYD 1540 1540 POTENTIAL.
FT CARBOHYD 1605 1605 POTENTIAL.
FT CARBOHYD 1763 1763 POTENTIAL.
FT CARBOHYD 1908 1908 POTENTIAL.
SQ SEQUENCE 2039 AA; 223589 MW; 666F9033 CRC32;

```

Query Match 64.0%; Score 32; DB 1; Length 2039;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 2 GGRKVF 7
    |||||
DB 200 GGRKVF 205

RESULT 12
DPD_BOVIN STANDARD; PRT: 1025 AA.
AC Q28007;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE DIHYDROXYRIMIDINE DEHYDROGENASE (NADP+) (EC 1.3.1.2) (DPD)
DE (DIHYDROURACIL DEHYDROGENASE) (DIHYDROTHYMINE DEHYDROGENASE)
  DPD.
GN BOS TAURUS (BOVINE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; ARTIODACTYLA.
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE: 97069948.
RA ALBIN N., JOHNSON M.R., DIASIO R.B.;
RL DNA SEQ. 6,243-250(1996).
CC -I- FUNCTION: INVOLVED IN PYRIMIDINE BASE DEGRADATION. CATALYZES THE
CC REDUCTION OF URACIL AND THYMINE.
CC -I- CATALYTIC ACTIVITY: 5,6-DIHYDROURACIL + NADP(+) = URACIL + NADPH.
CC -I- COFACTOR: FAD AND FEN FLAVOPROTEIN. ALSO CONTAINS TWO 4FE-4S
CC CLUSTERS.
CC -I- PATHWAY: INITIAL AND RATE-LIMITING ENZYME IN THE 3-STEP PATHWAY OF
CC URACIL AND THYMINE CATABOLISM AND IN THE PATHWAY LEADING TO THE
CC FORMATION OF BETA-ALANINE.
DR EMBL: U20981; G677951;
DR PROSITE: P800198; 4FR4S.FERREDOXIN: 1.
KW OXIDOREDUCTASE, NADP, FLAVOPROTEIN, FAD, IRON-SULFUR, 4FE-4S.
FT METAL 953 953 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
FT METAL 956 956 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
FT METAL 959 959 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
FT METAL 963 963 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).
FT METAL 986 986 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).
FT METAL 989 989 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).
FT METAL 992 992 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).
FT METAL 996 996 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).
SQ SEQUENCE 1025 AA; 111696 MW; 72CDE110 CRC32;

```

Query Match 64.0%; Score 32; DB 1; Length 1025;  
 Best Local Similarity 71.4%; Pred. No. 1.8e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

OY 1 CGRKVF 7
    |||||
DB 354 CGRKVF 360

```

```

RESULT 13
DPD_HUMAN STANDARD; PRT: 1025 AA.
AC Q12882; Q16694; Q16761;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE DIHYDROXYRIMIDINE DEHYDROGENASE (NADP+) PRECURSOR (EC 1.3.1.2) (DPD)
DE (DIHYDROURACIL DEHYDROGENASE) (DIHYDROTHYMINE DEHYDROGENASE).
GN DPD.
OC HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE: 94365020.
RA YOKOTA H., FERNANDEZ-SALGUERO P., FURUYA H., LIN K., MCBRIDE O.W.,
RA PODSCHUN B., SCHNACKERZ K.D., GONZALEZ F.J.;
RL J. BIOL. CHEM. 269:23192-23196(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA JOHNSON M.R., ALBIN N., SHAHINIAN H., DIASIO R.B.;
RA SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE OF 581-635 FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE: 97047101.
RA VREKEN P., VAN KUILENBURG A.B.P., MEISNA R., SMIT G.P.A.,
RA BAKKER H.D., DE ABREU R.A., VAN GENNIP A.H.;
RL J. INHERIT. METAB. DIS. 19:645-654(1996).
RN [4]
RP CHARACTERIZATION AND PARTIAL SEQUENCE.
RC TISSUE=LIVER;
RX MEDLINE: 92381021.
RA LU Z.-H., ZHANG R., DIASIO R.B.;

```

RL J. BIOL. CHEM. 267:17102-17109(1992).  
CC -1- FUNCTION: INVOLVED IN PYRIMIDINE BASE DEGRADATION. CATALYZES THE  
CC REDUCTION OF URACIL AND THYMINE. ALSO INVOLVED THE DEGRADATION OF  
CC THE CHEMOTHERAPEUTIC DRUG 5-FLUOROURACIL.  
CC -1- CATALYTIC ACTIVITY: 5,6-DIHYDROURACIL + NADP(+) = URACIL + NADPH.  
CC -1- COFACTOR: TWO EACH OF FAD AND FMN FLAVOPROTEIN. ALSO CONTAINS TWO  
CC 4FE-4S CLUSTERS. CONTAINS APPROXIMATELY 33 IRON ATOMS PER  
CC MOLECULE.  
CC -1- PATHWAY: INITIAL AND RATE-LIMITING ENZYME IN THE 3-STEP PATHWAY OF  
CC URACIL AND THYMINE CATABOLISM AND IN THE PATHWAY LEADING TO THE  
CC FORMATION OF BETA-ALANINE.  
CC -1- SUBUNIT: HOMODIMER.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -1- DISEASE: DEFECTS IN DPYD ARE THE CAUSE OF HEREDITARY THYMINE-  
CC URACILURIA (ALSO KNOWN AS FAMILIAL PYRIMIDINEMIA), A DISEASE  
CC CHARACTERIZED BY PERSISTENT URINARY EXCRETION OF EXCESSIVE AMOUNTS  
CC OF URACIL, THYMINE AND 5-HYDROXYMETHYLRACIL. PATIENTS SUFFERING  
CC FROM THIS DISEASE SHOW A SEVERE REACTION TO THE ANTICANCER DRUG 5-  
CC FLUOROURACIL. THIS REACTION INCLUDES STOMATITIS, LEUKOPENIA,  
CC THROMBOCYTOPENIA, HAIR LOSS, DIARRHEA, FEVER, MARKED WEIGHT LOSS,  
CC CEREBELLAR ATAXIA, AND NEUROLOGIC SYMPTOMS, PROGRESSING TO  
CC SEMICOMA.  
CC EMBL: U09178: G558305; -  
CC EMBL: U20938: G693912; -  
CC EMBL: X95670: E223407; -  
CC DR MIM: 274270; -  
CC DR PROSITE: PS00198: 4FE4S\_FERREDOXIN; 1.  
CC DR OXIDOREDUCTASE: NADP: FLAVOPROTEIN; FAD: IRON-SULFUR; 4FE-4S.  
CC FT PROPER 1 3  
CC FT CHAIN 1 1025 DIHYDROPYRIMIDINE DEHYDROGENASE.  
CC FT NP\_BIND 335 351 NADP (POTENTIAL).  
CC FT NP\_BIND 471 481 FAD (POTENTIAL).  
CC FT DOMAIN 661 678 URACIL-BINDING (POTENTIAL).  
CC FT METAL 953 953 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).  
CC FT METAL 956 956 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).  
CC FT METAL 959 959 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).  
CC FT METAL 963 963 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).  
CC FT METAL 986 986 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).  
CC FT METAL 989 989 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).  
CC FT METAL 992 992 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).  
CC FT METAL 996 996 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).  
CC FT CONFLICT 910 910 S -> N (IN REF. 2).  
CC FT SEQUENCE 1025 AA; 111374 MW; ED93DB7A CRC32;  
  
Query Match 64.0%; Score 32; DB 1; Length 1025;  
Best Local Similarity 71.4%; Pred. No. 1.8e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 CGGRKVF 7  
DB 354 CGARRVF 360  
  
RESULT 14  
DPYD\_PIG STANDARD; PRT: 1025 AA.  
ID DPYD\_PIG  
AC 028943;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DE DIHYDROPYRIMIDINE DEHYDROGENASE (NADP+) PRECURSOR (BC 1.3.1.2) (DPD)  
DE (DIHYDROURACIL DEHYDROGENASE) (DIHYDROTHYMINE DEHYDROGENASE).  
GN DPYD.  
OS SUS SCROFA (PIG).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC LUTHERIA; ARTIODACTYLA.  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE-LIVER:  
RX MEDLINE: 94365020.  
RA YOKOTA H., FERNANDEZ-SALGUERO P., FURUYA H., LIN K., MCBRIDE O.W.,  
RA PODSCHUN B., SCHNACKERZ K.D., GONZALEZ F.J.;

RL J. BIOL. CHEM. 269:23192-23196(1994).  
CC -1- FUNCTION: INVOLVED IN PYRIMIDINE BASE DEGRADATION. CATALYZES THE  
CC REDUCTION OF URACIL AND THYMINE.  
CC -1- CATALYTIC ACTIVITY: 5,6-DIHYDROURACIL + NADP(+) = URACIL + NADPH.  
CC -1- COFACTOR: FAD AND FMN FLAVOPROTEIN. ALSO CONTAINS TWO 4FE-4S  
CC CLUSTERS.  
CC -1- COFACTOR: TWO EACH OF FAD AND FMN FLAVOPROTEIN. ALSO CONTAINS TWO  
CC 4FE-4S CLUSTERS. CONTAINS APPROXIMATELY 30 IRON ATOMS PER  
CC MOLECULE.  
CC -1- PATHWAY: INITIAL AND RATE-LIMITING ENZYME IN THE 3-STEP PATHWAY OF  
CC URACIL AND THYMINE CATABOLISM AND IN THE PATHWAY LEADING TO THE  
CC FORMATION OF BETA-ALANINE.  
CC -1- SUBUNIT: HOMODIMER.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC EMBL: U09179: G558307; -  
CC DR PROSITE: PS00198: 4FE4S\_FERREDOXIN; 1.  
CC DR OXIDOREDUCTASE: NADP: FLAVOPROTEIN; FAD: IRON-SULFUR; 4FE-4S.  
CC FT PROPER 1 3  
CC FT CHAIN 1 1025 DIHYDROPYRIMIDINE DEHYDROGENASE.  
CC FT NP\_BIND 335 351 NADP (POTENTIAL).  
CC FT NP\_BIND 471 481 FAD (POTENTIAL).  
CC FT DOMAIN 661 678 URACIL-BINDING (POTENTIAL).  
CC FT METAL 953 953 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).  
CC FT METAL 956 956 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).  
CC FT METAL 959 959 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).  
CC FT METAL 963 963 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).  
CC FT METAL 986 986 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).  
CC FT METAL 989 989 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).  
CC FT METAL 992 992 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).  
CC FT METAL 996 996 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).  
CC FT SEQUENCE 1025 AA; 111424 MW; 0A297187 CRC32;  
  
Query Match 64.0%; Score 32; DB 1; Length 1025;  
Best Local Similarity 71.4%; Pred. No. 1.8e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 CGGRKVF 7  
DB 354 CGARRVF 360  
  
RESULT 15  
ENV\_HY1BN STANDARD; PRT: 852 AA.  
ID ENV\_HY1BN  
AC P12468;  
DT 01-OCT-1989 (REL. 12, CREATED)  
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)  
DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)  
DE ENVELOPE POLYPROTEIN GP160 PRECURSOR (CONTAINS: GP120 AND GP41).  
GN ENV.  
OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (BRAIN ISOLATE) (HIV-1).  
OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;  
OC LENTIVIRINAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 89085613.  
RA AMAND R., THAYER R., SRINIVASAN A., NAYYAR S., GARDNER M., LUCI P.,  
RA DANDEKAR S.;  
RL VIROLOGY 168:79-89(1989).  
CC -1- THIS VIRUS IS CYTOPATHICALLY ACTIVE AND WAS HARVESTED FROM THE  
CC BRAIN TISSUE OF A NEUROLOGICAL AIDS PATIENT.  
CC EMBL: M21098: G326430; -  
CC DR PIR: A31667; VCLDB.  
CC DR HIV: M21098; ENVSBRVA.  
CC AIDS; COAT PROTEIN; POLYPROTEIN; GLYCOPROTEIN; TRANSMEMBRANE;  
CC SIGNAL.  
FT SIGNAL 1 30  
FT CHAIN 31 507 EXTERIOR MEMBRANE GLYCOPROTEIN (GP120).  
FT CHAIN 508 852 TRANSMEMBRANE GLYCOPROTEIN (GP41).  
FT DISULFID 54 74 BY SIMILARITY.  
FT DISULFID 119 205 BY SIMILARITY.  
FT DISULFID 126 196 BY SIMILARITY.

FT	DISULFID	131	155	BY SIMILARITY.
FT	DISULFID	218	247	BY SIMILARITY.
FT	DISULFID	228	239	BY SIMILARITY.
FT	DISULFID	296	330	BY SIMILARITY.
FT	DISULFID	376	439	BY SIMILARITY.
FT	DISULFID	383	412	BY SIMILARITY.
FT	CARBOHYD	49	49	POTENTIAL.
FT	CARBOHYD	88	88	POTENTIAL.
FT	CARBOHYD	135	135	POTENTIAL.
FT	CARBOHYD	138	138	POTENTIAL.
FT	CARBOHYD	154	154	POTENTIAL.
FT	CARBOHYD	158	158	POTENTIAL.
FT	CARBOHYD	197	197	POTENTIAL.
FT	CARBOHYD	234	234	POTENTIAL.
FT	CARBOHYD	241	241	POTENTIAL.
FT	CARBOHYD	262	262	POTENTIAL.
FT	CARBOHYD	276	276	POTENTIAL.
FT	CARBOHYD	289	289	POTENTIAL.
FT	CARBOHYD	295	295	POTENTIAL.
FT	CARBOHYD	301	301	POTENTIAL.
FT	CARBOHYD	331	331	POTENTIAL.
FT	CARBOHYD	354	354	POTENTIAL.
FT	CARBOHYD	360	360	POTENTIAL.
FT	CARBOHYD	384	384	POTENTIAL.
FT	CARBOHYD	390	390	POTENTIAL.
FT	CARBOHYD	396	396	POTENTIAL.
FT	CARBOHYD	400	400	POTENTIAL.
FT	CARBOHYD	442	442	POTENTIAL.
FT	CARBOHYD	456	456	POTENTIAL.
FT	CARBOHYD	607	607	POTENTIAL.
FT	CARBOHYD	612	612	POTENTIAL.
FT	CARBOHYD	621	621	POTENTIAL.
FT	CARBOHYD	633	633	POTENTIAL.
FT	CARBOHYD	670	670	POTENTIAL.
FT	CARBOHYD	812	812	POTENTIAL.
SQ	SEQUENCE	852 AA;	97203 MW;	679CACDE CRC32;

Query Match 64.0%; Score 32; DB 1; Length 852;  
 Best Local Similarity 62.5%; Pred. No. 1.5e+02;  
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 CGGRKVC 8  
 DB 376 CGGEFFFC 383

Search completed: July 18, 1999, 11:37:53  
 Job time: 1752 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 06:02:51 ; Search time 63.5 seconds  
(without alignments)  
6.950 Million cell updates/sec

Title: US-09-142-043-7

Perfect score: 50

Sequence: 1 CGGRKYFC 8

Scoring table: BLOSUM62

Searched: 180763 segs, 55169189 residues

Database :

SPTREMBL-8:\*  
1: sp\_fungi:\*  
2: sp\_human:\*  
3: sp\_invertebrate:\*  
4: sp\_mammal:\*  
5: sp\_mhc:\*  
6: sp\_organelle:\*  
7: sp\_phage:\*  
8: sp\_plant:\*  
9: sp\_bacteria:\*  
10: sp\_fodent:\*  
11: sp\_virus:\*  
12: sp\_vertebrate:\*  
13: sp\_unclassified:\*  
14: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	37	74.0	140	8	023554	023554 arabidopsis
2	35	70.0	1698	3	094438	094438 chironomus
3	35	70.0	1704	3	094446	094446 chironomus
4	35	70.0	129	7	038289	038289 bacterioph
5	35	70.0	463	8	081025	081025 arabidopsis
6	35	70.0	119	11	071379	071379 human immun
7	34	68.0	4650	2	015598	015598 homo sapien
8	34	68.0	26926	2	010466	010466 homo sapien
9	34	68.0	1744	3	023425	023425 leishmania
10	34	68.0	340	8	080685	080685 arabidopsis
11	34	68.0	581	9	068924	068924 neisseria g
12	34	68.0	199	11	076471	076471 human immun
13	34	68.0	197	11	076474	076474 human immun
14	34	68.0	99	11	076485	076485 human immun
15	34	68.0	101	11	076525	076525 human immun
16	33	66.0	445	1	059684	059684 schizosacch
17	33	66.0	1077	1	074853	074853 schizosacch
18	33	66.0	172	3	045929	045929 caenorhabdi
19	33	66.0	582	8	065153	065153 chlamydomon
20	33	66.0	279	8	004110	004110 oryza sativ
21	33	66.0	436	9	083660	083660 bartonella
22	33	66.0	751	10	P70535	P70535 rattus norv
23	33	66.0	109	11	074753	074753 rattus norv
24	33	66.0	409	14	059500	059500 pyrococcus
25	33	66.0	343	14	029678	029678 archaeoglob
26	32	64.0	374	1	003289	003289 saccharomyc
27	32	64.0	2039	2	016745	016745 homo sapien
28	32	64.0	2489	2	016744	016744 homo sapien
29	32	64.0	444	3	018467	018467 caenorhabdi

## ALIGNMENTS

30	32	64.0	761	3	018468	018468 caenorhabdi
31	32	64.0	2165	3	019791	019791 caenorhabdi
32	32	64.0	661	4	029531	029531 pan troglod
33	32	64.0	2014	4	029530	029530 pan troglod
34	32	64.0	122	7	038249	038249 bacterioph
35	32	64.0	54	8	024576	024576 zea mays (m
36	32	64.0	76	9	P70850	P70850 borrelia bu
37	32	64.0	370	9	050184	050184 pseudomonas
38	32	64.0	432	9	052812	052812 amycolatops
39	32	64.0	295	9	056899	056899 yeisina en
40	32	64.0	1021	9	025200	025200 helicobacte
41	32	64.0	579	9	068016	068016 agrobacteri
42	32	64.0	1577	9	055265	055265 streptococc
43	32	64.0	269	9	055998	055998 synchococc
44	32	64.0	1025	10	089000	089000 rattus norv
45	32	64.0	115	11	039128	039128 human immun

RESULT 1  
ID 023554 PRELIMINARY; PRT: 140 AA.  
AC 023554;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 15.9 KD PROTEIN.  
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).  
OC EUKARYOTA; VIRIDIPHYTES; CHAROPHYTA/EMBRYOPHYTA GROUP; EMBRYOPHYTA;  
OC TRACHEOPHYTA; EUPHYLOPHYTES; SPERMATOPHYTES; MAGNOLIOPHYTA;  
OC EUDICOTYLEDONS; ROSIDAE; CAPPARIDAE; BRASSICACEAE; ARABIDOPSIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA BEVAN M., STIEKEMA W., MURPHY G., WAMBUTT R., POHL T., TERRYN N.,  
RA KREIS M., KAVANAGH T., ENTIAN K.D., RIEGER M., JAMES R.,  
RA PUIGDOMENECH P., HATZIOPOULOS P., OBERHAUSER B., DUESTERHOFT A.,  
RA JONES J., PALME K., ANSORGE W., DELSENY M., BANCROFT I., MEMES H.W.,  
RA SCHUELLER C., CHALWATZIS N.;  
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;  
RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: Z57342; E327039; -.  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 140 AA; 15935 MW; EA74EB96 CRC32;

Query Match 74.0%; Score 37; DB 8; Length 140;  
Best Local Similarity 85.7%; Pred. No. 7.6;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGGRKYFC 7  
DB 82 CGGQKVF 88  
RESULT 2  
ID 094438 PRELIMINARY; PRT: 1698 AA.  
AC 094438;  
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)  
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE 185 KDA SILK PROTEIN.  
GN SPI85.  
OS CHIRONOMUS PALUDIVITTATUS (MIDGE).  
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;  
OC PTERYGOTA; DIPTERA; NEMATOCERA; CHIRONOMIDEA; CHIRONOMIDAE;  
OC CHIRONOMINAE; CHIRONOMUS.  
RN [1]

RP SEQUENCE FROM N.A.  
RC TISSUE-SALIVARY GLAND;  
RA CASE S.T., COX C., BELL W.C., HOFFMAN R.T., MARTIN J., HAMILTON R.;  
RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: U54640; G1314732;  
SQ SEQUENCE 1698 AA; 186164 MW; 85BE8E520 CRC32;

Query Match 70.0%; Score 35; DB 3; Length 1698;  
Best Local Similarity 50.0%; Pred. No. 1.4e+02;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 CGGRKVC 8  
Db 1532 CGAKKWC 1539

RESULT 3  
ID 094446 PRELIMINARY; PRT; 1704 AA.  
AC 094446;

DT 01-FEB-1997 (TREMBLREL. 02, CREATED)  
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE 220 KDA SILK PROTEIN.

GN SP220.  
OS CHIRONOMUS THUMMI THUMMI (MIDGE).  
OC EUARCYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;  
CC PERKYOTA; DIPTERA; NEMATOCERA; CHIRONOMIDEA; CHIRONOMIDAE;  
CC CHIRONOMINAE; CHIRONOMUS.

RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-SALIVARY GLAND;  
RA CASE S.T., COX C., BELL W.C., HOFFMAN R.T., MARTIN J., HAMILTON R.;  
RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: U54641; G1314734;  
SQ SEQUENCE 1704 AA; 185745 MW; 0245A38E CRC32;

Query Match 70.0%; Score 35; DB 3; Length 1704;  
Best Local Similarity 75.0%; Pred. No. 1.4e+02;

Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 CGGRKVC 8  
Db 1540 CGEKKVC 1547

RESULT 4  
ID 038289 PRELIMINARY; PRT; 129 AA.  
AC 038289;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE E8.

GN E8.  
OS BACTERIOPHAGE C2.  
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; SIPHOVIRIDAE.  
RN [1]  
RP SEQUENCE FROM N.A.

RA MEDLINE: 94036453.  
RA WARD L.J., BERSFORD T.P., LUBBERS M.W., JARVIS B.D., JARVIS A.W.;  
RT "Sequence analysis of the lysin gene region of the prolate lactococcal  
bacteriophage C2.";  
RL CAN. J. MICROBIOL. 39:767-774(1993).

RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 95115663.  
RA LUBBERS M.W., WARD L.J., BERSFORD T.P., JARVIS B.D., JARVIS A.W.;  
RT "Sequencing and analysis of the cos region of the lactococcal  
bacteriophage C2.";  
RL MOL. GEN. GENET. 245:160-166(1994).

RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 96086019.  
RA LUBBERS M.W., WATERFIELD N.R., BERSFORD T.P., LE PAGE R.W.,  
RA JARVIS A.W.;

RT "Sequencing and analysis of the prolate-headed lactococcal  
bacteriophage C2 genome and identification of the structural genes.";  
RL APPL. ENVIRON. MICROBIOL. 61:4348-4356(1995).  
DR EMBL: L48605; G1146291;  
SQ SEQUENCE 129 AA; 15434 MW; A26F8FCC CRC32;

Query Match 70.0%; Score 35; DB 7; Length 129;  
Best Local Similarity 62.5%; Pred. No. 16;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 CGGRKVC 8  
Db 106 CNARKVC 113

RESULT 5  
ID 081025 PRELIMINARY; PRT; 463 AA.  
AC 081025;

DT 01-NOV-1998 (TREMBLREL. 08, CREATED)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE PUTATIVE AIG1 PROTEIN.

GN F12C20.14.  
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).  
OC EUKARYOTA; VIRIDIPHYTES; CHAROPHYTA/EMBRYOPHYTA GROUP; EMBRYOPHYTA;  
CC TRACHEOPHYTA; EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA;  
CC EUDICOTYLEDONS; ROSIDAE; CAPRIFALES; BRASSICACEAE; ARABIDOPSIS.

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV, COLUMBIA;  
RA ROUNDSLEY S.D., RONNING C.M., LIN X., KETCHUM K.A., CROSBY M.L.,  
RA BRANDON R.C., SYKES S.M., KAUL S., MASON T.M., KERLAVAGE A.R.,  
RA ADAMS M.D., SOMERVILLE C.R., VENTER J.C.;

RT "Arabidopsis thaliana chromosome II BAC F12C20 genomic sequence.";  
RL SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: AC005168; G3426044;  
SQ SEQUENCE 463 AA; 52217 MW; 1B10E466 CRC32;

Query Match 70.0%; Score 35; DB 8; Length 463;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGGRKVC 6  
Db 162 CGGRKVC 167

RESULT 6  
ID 071379 PRELIMINARY; PRT; 119 AA.  
AC 071379;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE ENVELOPE GLYCOPROTEIN, C2-V3 REGION (FRAGMENT).

GN ENV.  
OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).  
OC VIRUSES; RETROID VIRUSES; RETROVIRIDAE; LENTIVIRUS.  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN-CR202;  
RA SHAO Y., WOLF H.;  
RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.

RC STRAIN-CR202;  
RA BRYANT B.W.;  
RL SUBMITTED (JAN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: U20053; G901976;  
DR PFAM: PF00516; GP120; 1.  
KW ENVELOPE PROTEIN.  
FT NON\_TER 1 1  
FT NON\_TER 119 119  
SQ SEQUENCE 119 AA; 13190 MW; 264A3A1C CRC32;

Query Match 70.0%; Score 35; DB 11; Length 119;  
Best Local Similarity 62.5%; Pred. No. 15;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 CGGRKVC 8  
DB 112 CGGESFEC 119

RESULT 7  
O15598 PRELIMINARY; PRT; 4650 AA.  
AC 015598;  
DT 01-NOV-1996 (TREMELREL. 01, CREATED)  
DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMELREL. 08, LAST ANNOTATION UPDATE)  
DE TITIN (FRAGMENT).  
GN TITIN.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-HEART;  
RX MEDLINE: 94008990.  
RA GAUTEL M., LEONARD K., LABELT S.;  
RT "Phosphorylation of KSP motifs in the C-terminal region of titin in  
RT differentiating myoblasts."  
DR EMBL: X69490; G407139;  
DR PFAM: PF00041; fn3; 16.  
DR PFAM: PF00047; 19; 12.  
DR PFAM: PF00069; kinase; 1.  
FT NON\_TER 1 1  
SQ SEQUENCE 4650 AA; 522065 MW; 1D2AD342 CRC32;

Query Match 68.0%; Score 34; DB 2; Length 4650;  
Best Local Similarity 83.3%; Pred. No. 5.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGGRKV 6  
DB 4591 CGGRKI 4596

RESULT 8  
O10466 PRELIMINARY; PRT; 26926 AA.  
AC 010466;  
DT 01-NOV-1996 (TREMELREL. 01, CREATED)  
DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMELREL. 08, LAST ANNOTATION UPDATE)  
DE TITIN, HEART ISOFORM N2-B (EC 2.7.1.-) (CONNECTIN).  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-HEART;  
RX MEDLINE: 96026330.  
RA LABELT S., KOLMER B.;

RT "Titins: giant proteins in charge of muscle ultrastructure and  
RT elasticity."  
RL SCIENCE 270:293-296(1995).  
RN [2]  
RP SEQUENCE OF 22277-25376 FROM N.A.  
RX MEDLINE: 92258380.  
RA LABELT S., GAUTEL M., LAKEY A., TRINICK J.;  
RT "Towards a molecular understanding of titin."  
RL EMBL J. 11:1711-1716(1992).  
RN [3]

RP SEQUENCE OF 1976-2014 FROM N.A.  
RA LABELT S.;  
RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [4]

RP CHARACTERIZATION.  
RX MEDLINE: 95331314.  
RA GAUTEL M., CASTIGLIONE-MORELLI M.A., PFUHL M., MOTTA A., PASTORE A.;  
RT "A calmodulin-binding sequence in the C-terminus of human cardiac  
RT titin kinase."  
RL EUR. J. BIOCHEM. 230:752-759(1995).  
CC -1- FUNCTION: THIS GIANT MUSCLE PROTEIN MAY BE INVOLVED IN MUSCLE  
CC ASSEMBLY AND IN MAINTAINING THE STRUCTURAL INTEGRITY OF  
CC SARCOMERES. MAY HAVE PROTEIN KINASE ACTIVITY.  
CC -1- TISSUE SPECIFICITY: MUSCLE-SPECIFIC.  
CC -1- ALTERNATIVE PRODUCTS: A NUMBER OF FORMS OF THIS PROTEIN ARE  
CC PRODUCED BY ALTERNATIVE SPLICING WHICH DIFFER IN TISSUE  
CC DISTRIBUTION. DIFFERENT SIZE TRANSCRIPTS MAY ALSO EXIST WITHIN ANY  
CC ONE TISSUE. THE SEQUENCE SHOWN HERE IS THAT OF THE HEART ISOFORM  
CC N2-B.  
CC -1- SIMILARITY: TO THE CATALYTIC DOMAINS OF OTHER SERINE/THREONINE  
CC KINASES.  
CC -1- SIMILARITY: BELONGS TO IMMUNOGLOBULIN SUPERFAMILY. CONTAINS 112  
CC IMMUNOGLOBULIN C2-LIKE DOMAINS AND 132 FIBRONECTIN TYPE IIT-LIKE  
CC DOMAINS.  
DR EMBL: X64698; G37193;  
DR EMBL: X83270; G602580;  
DR EMBL: X64697; G37195;  
DR EMBL: X90568; G1017425;  
DR EMBL: X64699; G37191;  
DR PFAM: PF00041; fn3; 132.  
DR PFAM: PF00047; 19; 58.  
DR PFAM: PF00069; kinase; 1.  
KW MUSCLE PROTEIN; CYTOSKELETON; STRUCTURAL PROTEIN; CALMODULIN-BINDING;  
KW SERINE/THREONINE-PROTEIN KINASE; ALTERNATIVE SPLICING; REPEAT;  
KW IMMUNOGLOBULIN FOLD; PHOSPHORYLATION.  
FT DOMAIN 1370 1389  
FT DOMAIN 4429 4614  
FT DOMAIN 24731 25070  
FT DOMAIN 25030 25056  
FT MOD\_RES 1372 1372  
FT MOD\_RES 1377 1377  
FT MOD\_RES 1382 1382  
FT MOD\_RES 1387 1387  
FT MOD\_RES 26171 26171  
FT MOD\_RES 26178 26178  
FT MOD\_RES 26184 26184  
FT MOD\_RES 26190 26190  
FT MOD\_RES 22277 22277  
FT CONFLICT 22449 22449  
FT CONFLICT 22454 22454  
FT CONFLICT 23324 23324  
SQ SEQUENCE 26926 AA; 2993428 MW; 0214E3A3 CRC32;

Query Match 68.0%; Score 34; DB 2; Length 26926;  
Best Local Similarity 83.3%; Pred. No. 2.2e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGGRKV 6  
DB 26867 CGGRKI 26872

```
RESULT 9
DE RECD.
GN RECD.
OS NEISSERIA GONORRHOEA.
OC BACTERIA; PROTEOBACTERIA; BETA SUBDIVISION; NEISSERIACEAE; NEISSERIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MS11;
RA CHAUSSEE M.S., WILSON J., HILL S.A.;
RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF058330; G3063663; -.
SO SEQUENCE 581 AA; 63039 MW; 66AFE708 CRC32;

Query Match
Best Local Similarity 68.0%; Score 34; DB 9; Length 581;
Pred. No. 89;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGRKVC 8
DB 551 GGRKVC 557

RESULT 12
DE RECD.
GN RECD.
OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).
OC VIRUSES; RETROVIRIDAE; LENTIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95018660.
RA KOSBER B.T., KUNSTMAN K.J., PATTERSON B.K., FURTADO M., MCEVILLY M.M.,
RA LEVY R., WOLINSKY S.M.;
RT "Genetic differences between blood- and brain-derived viral sequences
RT of conserved elements in the V3 region of the envelope protein of
RT brain-derived sequences."
RL J. VIROL. 68:7467-7481(1994).
DR EMBL; U05564; G607621; -.
DR PFAM; PF00516; GP120; 1.
KW ENVELOPE PROTEIN.
FT NON_TER 1
FT NON_TER 199
FT NON_TER 199
SQ SEQUENCE 199 AA; 21975 MW; 3F8C32A5 CRC32;

Query Match
Best Local Similarity 68.0%; Score 34; DB 11; Length 199;
Pred. No. 36;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGGRKVC 8
DB 102 CGGRKVC 109

RESULT 13
DE RECD.
GN RECD.
OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).
OC VIRUSES; RETROVIRIDAE; LENTIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95018660.
RA KOSBER B.T., KUNSTMAN K.J., PATTERSON B.K., FURTADO M., MCEVILLY M.M.,
RA LEVY R., WOLINSKY S.M.;
RT "Genetic differences between blood- and brain-derived viral sequences
RT of conserved elements in the V3 region of the envelope protein of
RT brain-derived sequences."
RL J. VIROL. 68:7467-7481(1994).
DR EMBL; U05564; G607621; -.
DR PFAM; PF00516; GP120; 1.
KW ENVELOPE PROTEIN.
FT NON_TER 1
FT NON_TER 199
FT NON_TER 199
SQ SEQUENCE 199 AA; 21975 MW; 3F8C32A5 CRC32;

Query Match
Best Local Similarity 68.0%; Score 34; DB 8; Length 340;
Pred. No. 56;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGGRKVC 7
DB 45 CGGRKVC 51

RESULT 11
DE RECD.
GN RECD.
OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).
OC VIRUSES; RETROVIRIDAE; LENTIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95018660.
RA KOSBER B.T., KUNSTMAN K.J., PATTERSON B.K., FURTADO M., MCEVILLY M.M.,
RA LEVY R., WOLINSKY S.M.;
RT "Genetic differences between blood- and brain-derived viral sequences
RT of conserved elements in the V3 region of the envelope protein of
RT brain-derived sequences."
RL J. VIROL. 68:7467-7481(1994).
DR EMBL; U05564; G607621; -.
DR PFAM; PF00516; GP120; 1.
KW ENVELOPE PROTEIN.
FT NON_TER 1
FT NON_TER 199
FT NON_TER 199
SQ SEQUENCE 199 AA; 21975 MW; 3F8C32A5 CRC32;

Query Match
Best Local Similarity 68.0%; Score 34; DB 3; Length 1724;
Pred. No. 2,2e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGGRKVC 8
DB 1479 CGGRKVC 1486

RESULT 10
DE RECD.
GN RECD.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPHYTES; CHAROPHYTA/EMBRYOPHYTA GROUP; EMBRYOPHYTA;
OC TRACHEOPHYTA; EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA;
OC EUDICOTYLEDONS; ROSIDAE; CAPRALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-CV, COLUMBIA.
RA ROUNDSLEY S.D., KAUL S., LIN X., KETCHUM K.A., CROSBY M.L.,
RA BRANDON R.C., SYKES S.M., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA SOMERVILLE C.R., VENTER J.C.;
RT "Arabidopsis thaliana chromosome II BAC T3K9 genomic sequence."
RL SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AC004261; G3402718; -.
SQ SEQUENCE 340 AA; 38701 MW; 742BPA49 CRC32;
```

```
RESULT 9
DE RECD.
GN RECD.
OS NEISSERIA GONORRHOEA.
OC BACTERIA; PROTEOBACTERIA; BETA SUBDIVISION; NEISSERIACEAE; NEISSERIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TAR11;
RA LEGARE D., HETTEMER E., OUELLETTE M.;
RT "The P-glycoprotein-related gene family in Leishmania."
RL MOI. BIOCHEM. PARASITOL. 68:81-91(1994).
DR EMBL; L29485; G460313; -.
DR PFAM; PF00005; ABC_tran; 2.
DR PFAM; PF00664; ABC_membrane; 2.
SQ SEQUENCE 1724 AA; 188967 MW; C9F31C24 CRC32;

Query Match
Best Local Similarity 62.5%; Pred. No. 2,2e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGGRKVC 8
DB 1479 CGGRKVC 1486

RESULT 10
DE RECD.
GN RECD.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPHYTES; CHAROPHYTA/EMBRYOPHYTA GROUP; EMBRYOPHYTA;
OC TRACHEOPHYTA; EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA;
OC EUDICOTYLEDONS; ROSIDAE; CAPRALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-CV, COLUMBIA.
RA ROUNDSLEY S.D., KAUL S., LIN X., KETCHUM K.A., CROSBY M.L.,
RA BRANDON R.C., SYKES S.M., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA SOMERVILLE C.R., VENTER J.C.;
RT "Arabidopsis thaliana chromosome II BAC T3K9 genomic sequence."
RL SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AC004261; G3402718; -.
SQ SEQUENCE 340 AA; 38701 MW; 742BPA49 CRC32;
```

RX MEDLINE: 95018660.  
 RA KOBBER B.T., KUNSTMAN K.J., PATTERSON B.K., FURTADO M., MCEVILLY M.M.,  
 RA LEVY R., WOLINSKY S.M.;  
 RT "Genetic differences between blood- and brain-derived viral sequences  
 from human immunodeficiency virus type 1-infected patients: evidence  
 of conserved elements in the V3 region of the envelope protein of  
 brain-derived sequences.";  
 RL J. VIROL. 68:7467-7481(1994).  
 DR EMBL: U05567; G607627; -.  
 DR PFAM: PF00516; GP120; 1.  
 KW ENVELOPE PROTEIN.  
 FT NON\_TER 1 1  
 FT NON\_TER 197 197  
 SQ SEQUENCE 197 AA; 21953 MW; 17171ED5 CRC32;

Query Match 68.0%; Score 34; DB 11; Length 197;  
 Best Local Similarity 62.5%; Pred. No. 36;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 CGGRKVC 8  
 DB 101 CGGEPLFC 108

RESULT 14

ID 076485 PRELIMINARY; PRT; 99 AA.  
 AC 076485;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DE ENVELOPE GLYCOPROTEIN V4V5 REGION (FRAGMENT).  
 GN ENV.  
 OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).  
 OC VIRUSES; RETROID VIRUSES; RETROVIRIDAE; LENTIVIRUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 95018660.  
 RA KOBBER B.T., KUNSTMAN K.J., PATTERSON B.K., FURTADO M., MCEVILLY M.M.,  
 RA LEVY R., WOLINSKY S.M.;  
 RT "Genetic differences between blood- and brain-derived viral sequences  
 from human immunodeficiency virus type 1-infected patients: evidence  
 of conserved elements in the V3 region of the envelope protein of  
 brain-derived sequences.";  
 RL J. VIROL. 68:7467-7481(1994).  
 DR EMBL: U05488; G607469; -.  
 DR PFAM: PF00516; GP120; 1.  
 KW ENVELOPE PROTEIN.  
 FT NON\_TER 1 1  
 FT NON\_TER 99 99  
 SQ SEQUENCE 99 AA; 10999 MW; 7C97F8BD CRC32;

Query Match 68.0%; Score 34; DB 11; Length 99;  
 Best Local Similarity 50.0%; Pred. No. 20;  
 Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 CGGRKVC 8  
 DB 6 CGGESLYC 13

RESULT 15

ID 076525 PRELIMINARY; PRT; 101 AA.  
 AC 076525;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DE ENVELOPE GLYCOPROTEIN V4V5 REGION (FRAGMENT).  
 GN ENV.  
 OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).

OC VIRUSES; RETROID VIRUSES; RETROVIRIDAE; LENTIVIRUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 95018660.  
 RA KOBBER B.T., KUNSTMAN K.J., PATTERSON B.K., FURTADO M., MCEVILLY M.M.,  
 RA LEVY R., WOLINSKY S.M.;  
 RT "Genetic differences between blood- and brain-derived viral sequences  
 from human immunodeficiency virus type 1-infected patients: evidence  
 of conserved elements in the V3 region of the envelope protein of  
 brain-derived sequences.";  
 RL J. VIROL. 68:7467-7481(1994).  
 DR EMBL: U05509; G607511; -.  
 DR PFAM: PF00516; GP120; 1.  
 KW ENVELOPE PROTEIN.  
 FT NON\_TER 1 1  
 FT NON\_TER 101 101  
 SQ SEQUENCE 101 AA; 11137 MW; 96C207B7 CRC32;

Query Match 68.0%; Score 34; DB 11; Length 101;  
 Best Local Similarity 62.5%; Pred. No. 20;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 CGGRKVC 8  
 DB 6 CGGEPLFC 108

Search completed: July 18, 1999, 06:02:53  
 Job time: 3108 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 00:42:55 ; Search time 58.51 Seconds  
(without alignments)  
7.950 Million cell updates/sec

Title: US-09-142-043-8

Perfect score: 123

Sequence: 1 FELVGEPSITYSTNDQVGIMSG 23

Scoring table: BLOSUM62

Searched: 162890 segs, 20225328 residues

Database: A\_Geneseq\_34.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	123	100.0	23	1 W31838	Peptide 4 from the
2	118	95.9	2317	1 P92219	CRI protein New nu
3	118	95.9	2039	1 R11810	Human complement t
4	118	95.9	1537	1 R11982	Partial human comp
5	118	95.9	543	1 R28543	CRI-4 (35E, 37Y) a
6	118	95.9	543	1 R28544	CRI-4 (35E) analog
7	118	95.9	543	1 R28545	CRI-4 (37Y) analog
8	118	95.9	543	1 R28546	CRI-4 (44T, 47D, 4
9	118	95.9	543	1 R28547	CRI-4 (52S, 53S, 5
10	118	95.9	543	1 R28548	CRI-4 (57V, 59K) a
11	118	95.9	543	1 R28549	CRI-4 (64K, 65T) a
12	118	95.9	543	1 R28550	CRI-4 (64K) analog
13	118	95.9	543	1 R28551	CRI-4 (65T) analog
14	118	95.9	543	1 R28552	CRI-4 (78T, 79D) a
15	118	95.9	543	1 R28553	CRI-4 (85R, 87N) a
16	118	95.9	543	1 R28554	CRI-4 (92T, 94H) a
17	118	95.9	543	1 R28555	CRI-4 (92T) analog
18	118	95.9	543	1 R28556	CRI-4 (94H) analog
19	118	95.9	543	1 R28557	CRI-4 (99H, 103E)
20	118	95.9	543	1 R28558	CRI-4 (109N, 110A,
21	118	95.9	543	1 R28559	CRI-4 (114-117STR
22	118	95.9	543	1 R28560	CRI-4 (114S) analo
23	118	95.9	543	1 R28561	CRI-4 (115T) analo
24	118	95.9	543	1 R28562	CRI-4 (116K) analo
25	118	95.9	543	1 R28563	CRI-4 (117P) analo
26	118	95.9	543	1 R28564	CRI-4 (116K, 117P)
27	118	95.9	543	1 R28565	CRI-4 (121Q) analo
28	118	95.9	483	1 R29091	CRI-4 (amino acids
29	118	95.9	481	1 R29092	CRI-4 (amino acids
30	118	95.9	543	1 R28566	CRI-4 (318R, 319N)
31	118	95.9	543	1 R28567	CRI-4 (318-321 RNP
32	118	95.9	543	1 R28568	CRI-4 (347T, 349Y)
33	118	95.9	543	1 R28569	CRI-4 (369-376 STR
34	118	95.9	543	1 R28570	CRI-4 (266-274 KKK
35	118	95.9	543	1 R28571	CRI-4 (364-367 NAA
36	118	95.9	2039	1 R36743	CRI. Nucleic acid
37	118	95.9	197	1 R47152	Sequence of solubi
38	118	95.9	76	1 R47153	Sequence of solubi
39	118	95.9	254	1 R47154	Sequence of solubi
40	118	95.9	133	1 R47155	Sequence of solubi
41	118	95.9	211	1 R47156	Sequence of solubi
42	118	95.9	209	1 W45909	SCR 1-3 of complen
43	118	95.9	209	1 W45912	SCR 1-3 of complen

## ALIGNMENTS

44 118 95.9 198 1 W45897 SCR 1-3 of complen  
45 118 95.9 1930 1 W45899 Human complement r

RESULT 1  
W31838  
ID W31838 standard; peptide: 23 AA.  
AC W31838; 26-MAR-1998 (first entry)  
DE Peptide 4 from the short consensus repeat 3 of complement receptor 1.  
KW Short consensus repeat 3; long homologous repeat A; LHR-A;  
KW complement receptor 1; CRI: complement inhibition;  
KW anti-haemolytic activity; inflammation; thrombotic condition;  
KW inappropriate complement activation; ARDS; Alzheimer's disease.  
OS Synthetic.  
OS Homo sapiens.  
PN WO9731944-A1.  
PD 04-SEP-1997.  
PF 26-FEB-1997; E00994.  
PR 02-MAR-1996; GB-004518.  
PA (SMK ) SMITHKLINE BEECHAM PLC.  
PI Edge CM, Mossakowska DEL, Smith RAG;  
DR MPI: 97-448630/41.  
PT Peptide derived from short consensus repeat 3 of human complement  
PT receptor 1 - between amino acids Cys154-Gly186, useful to treat  
PT disorder or disease associated with inflammation or inappropriate  
PT complement activation  
PS Claim 13; Page 25; 33pp; English.  
CC Synthetic soluble peptides W31835-38 are derived from the short  
CC consensus repeat 3 of the long homologous repeat A (LHR-A) of human  
CC complement receptor 1 (CRI). The present peptide is located between  
CC residues 164 and 186. These peptides have functional complement  
CC inhibitory, including anti-haemolytic, activity. Enhanced activity  
CC may be achieved by linking the peptides to a core structure  
CC (e.g. the MAP peptide) to produce a multimeric or chimeric polypeptide.  
CC This polypeptide can be used to treat a disorder or disease associated  
CC with inflammation or inappropriate complement activation. It can also be  
CC used to treat a thrombotic condition, adult respiratory distress syndrome  
CC (ARDS), wounds, Alzheimer's disease or a CNS inflammatory disorder, or  
CC delay hyperacute allograft or xenograft rejection.  
SQ Sequence 23 AA;

Query Match 100.0%; Score 123; DB 1; Length 23;  
Best Local Similarity 100.0%; Pred. No. 3e-13;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 FELVGEPSITYSTNDQVGIMSG 23  
DB 1 FELVGEPSITYSTNDQVGIMSG 23  
RESULT 2  
ID P92219  
AC P92219; P92219 standard; protein: 2317 AA.  
DT 22-FEB-1990 (first entry)  
DE CRI protein  
KW Complement: cofactor.  
OS Homo sapiens (human).  
EH Key Location/Qualifiers  
FT peptide 10..50  
FT /label= signal\_peptide  
PN WO8909220-A.  
PD 05-OCT-1989.  
PF 31-MAR-1989; U01358.  
PR 01-APR-1988; US-176532.  
PA (TCEI) T Cell Sciences Inc; (UYJO) The Johns Hopkins University;  
PA (BRIG\*) The Brigham and Women's Hospital.  
PI Fearon DT, Klickstein LB, Wong W, Carson G, Concino MF, Makrides SC;

DR WPI: 89-309498/42.  
DR N-PSDB: N91477.  
PT New nucleic acid sequences encoding new CRI protein - and its fragment,  
PT for diagnosis and control of complement-related immune defects,  
PT inflammation, myocardial infarct, etc  
PS Claim 1; fig. 1; 19pp; English.  
CC This is full-length CRI protein, and shortened forms are new, lacking  
CC the transmembrane region. The proteins and fragments bind C3b and/or  
CC C4b, have cofactor activity and inhibit C3 and C5 convertase activity.  
CC In the sequence, x-untranslated region. This has 7 short consensus  
CC repeats (CR) of 4 long direct homologous repeats (LHR) (see fig. 10 in  
CC patent). Deletion mutants, eg LHR-A for binding of C4b, and LHR-C  
CC for C3b binding, be made, and LHR-B and -C can detect CR2 sequences.  
CC They are useful in diagnosing and treating immune disorders, and prevent  
CC perfusion injury.  
SQ Sequence 2317 AA;

Query Match 95.9%; Score 118; DB 1; Length 2317;  
Best Local Similarity 95.7%; Pred. No. 3.6e-10;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FELVGEPSITSTNDQVGIMSG 23  
DB 214 FELVGEPSITCTSDQVGIMSG 236

RESULT 3  
ID R11810 standard; Protein; 2039 AA.  
AC R11810.  
DT 25-JUN-1991 (first entry)  
DE Human complement type 1 receptor.  
KW Complement system: C3b/C4b receptor; CRI; allergic reaction;  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT peptide 1..41  
FT /label- putative signal peptide  
FT 42..2039  
FT /label- CRI  
FT protein  
PN WO9105047-A.  
PD 18-APR-1991.  
PR 25-SEP-1990; U05454.  
PR 26-SEP-1989; US-412745.  
PR 26-SEP-1990; US-912349.  
PA (TCELL-) T CELL SCI INC.  
PA (UYJO ) JOHNS HOPKINS UNIVERSITY.  
PA (BRIG-) BRIGHAM AND WOMEN'S HOSPITAL.  
PI Fearon DT, Klickstein LB, Wong WW, Carson GR, Hoh M, Concino MF,  
PI Makrides SC, Marsh HC;  
DR WPI: 91-132854/18.  
DR N-PSDB: Q11642.  
PT Human complement receptor type 1 gene, encoded proteins and  
PT fragments - for treatment of immune disorders, myocardial infarct,  
PT damage due to inflammation and in treatment of thrombosis  
PS Claim 41; Fig 1; 234pp; English.  
CC The invention also covers fragments of this protein which have the  
CC ability to bind C3b and/or C4b, have cofactor I activity or can  
CC inhibit C3 or C5 convertase activity. The full-length protein, or  
CC its specified fragments are used to treat patients with immune  
CC disorders or a disorder caused by inappropriate complement  
CC activity. The protein is also used to treat thrombotic conditions  
CC in humans and animals. See also Q11643.  
SQ Sequence 2039 AA;

Query Match 95.9%; Score 118; DB 1; Length 2039;  
Best Local Similarity 95.7%; Pred. No. 3.1e-10;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FELVGEPSITSTNDQVGIMSG 23  
FH Key Location/Qualifiers

DB 205 FELVGEPSITCTSDQVGIMSG 227

RESULT 4  
ID R11982 standard; Protein; 1537 AA.  
AC R11982.  
DT 25-JUN-1991 (first entry)  
DE Partial human complement type 1 receptor.  
KW Complement system: C3b/C4b receptor; CRI; allergic reaction;  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT region 1..438  
FT /label- LHR-B  
FT 439..891  
FT /label- LHR-C  
FT region 892..1341  
FT /label- LHR-D  
FT region 1495..1498  
FT /note- "positively-charged; preceded by hydrophobic  
FT sequence"  
FT region 1521..1526  
FT /note- "has 67 per cent homology to site of protein  
FT kinase C phosphorylation in the EGF  
FT receptor"

WO9105047-A.  
PD 18-APR-1991.  
PR 25-SEP-1990; U05454.  
PR 26-SEP-1989; US-412745.  
PR 26-SEP-1990; US-912349.  
PA (TCELL-) T CELL SCI INC.  
PA (UYJO ) JOHNS HOPKINS UNIVERSITY.  
PA (BRIG-) BRIGHAM AND WOMEN'S HOSPITAL.  
PI Fearon DT, Klickstein LB, Wong WW, Carson GR, Hoh M, Concino MF,  
PI Makrides SC, Marsh HC;  
DR WPI: 91-132854/18.  
DR N-PSDB: Q11643.  
PT Human complement receptor type 1 gene, encoded proteins and  
PT fragments - for treatment of immune disorders, myocardial infarct,  
PT damage due to inflammation and in treatment of thrombosis  
PS Disclosure: Fig 5; 234pp; English.  
CC This sequence comprises three of the four tandem, direct, long  
CC homologous repeats of the full-length F allozyme of CRI. LHR-A is  
CC absent. Each LHR might represent a single C3b/C4b binding domain,  
CC making the receptor multivalent. The LHR's are composed of 7 short  
CC consensus repeats of 60-70 residues resembling the SCR's of other  
CC C3/C4 binding proteins. The protein and fragments of it having C3b  
CC and/or C4b binding activity can be used to treat immune disorders  
CC or disorders involving inappropriate complement activity.  
SQ Sequence 1537 AA;

Query Match 95.9%; Score 118; DB 1; Length 1537;  
Best Local Similarity 95.7%; Pred. No. 2.2e-10;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FELVGEPSITSTNDQVGIMSG 23  
DB 153 FELVGEPSITCTSDQVGIMSG 175

RESULT 5  
ID R28543 standard; peptide; 543 AA.  
AC R28543;  
DT 19-MAR-1993 (first entry)  
DE CRI-4 (3SE, 37Y) analogue.  
KW short consensus repeat; regulator of complement activation;  
KW C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.  
FH Key Location/Qualifiers



```
FT region 1.60
FT /label= SCR-1
FT region 61..122
FT /label= SCR-2
FT region 451..510
FT /label= SCR-8
FT region 511..543
FT /label= SCR-9
FT /note="TRUNCATED"
FT misc_difference 35
FT /note="Gly substituted by Glu (SCR-8)"
FT /note="Ser substituted by Tyr (SCR-8)"
FT EP-512733-A.
FT 11-NOV-1992.
FT 28-APR-1992: 303826.
FT 03-MAY-1991: US-695514.
FT (UNIW ) UNIV WASHINGTON.
FT Atkinson JP, Hourcade D, Kiyoch M;
FT MPI; 92-375009/46.
FT Complement activity regulator protein analogues - useful for
FT treating auto:immune diseases, to suppress transplant rejection,
FT for diagnosis etc.
FT Claim 11; Fig 2 and R11810; 23pp; English.
FT The cDNA clone designated CRI-4 was described in J. Exp. Med. (1988)
FT 168:1255-1270. It encodes the first 8 and a half amino terminal
FT SCRs of CRI. The invention concerns analogues of "regulator of
FT complement activation" proteins or truncated, hybrid or recombinant
FT forms of them. CRI-4 is a preferred truncated form and a number of
FT specified substitution variants of it are claimed. Positions 35 and
FT 37 of SCR-1 and the corresponding positions in SCR-8 have been
FT identified as important in C4b binding. The specification does not
FT contain the CRI-4 sequence; the sequence given here was constructed
FT from the full-length CRI amino acid sequence having GENSESQ
FT accession number R11810 and descriptions in the disclosure.
FT Sequence 543 AA;

SQ
Query Match 95.9%; Score 118; DB 1; Length 543;
Best Local Similarity 95.7%; Pred. No. 6.8e-11;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FELVGEPSIVTSNDQVGIMSG 23
Db 164 FELVGEPSIVCTSNDDQVGIMSG 186

RESULT 6
R28544 ID R28544 standard; peptide: 543 AA.
AC R28544:
DT 19-MAR-1993 (first entry)
DE CRI-4 (35E) analogue.
KW Short consensus repeat; regulator of complement activation;
KW C3b binding; C4b binding; human complement type 1 receptor.
OS Homo sapiens.
PI Key Location/Qualifiers
FH 1.60
FT /label= SCR-1
FT region 61..122
FT /label= SCR-2
FT region 451..510
FT /label= SCR-8
FT region 511..543
FT /label= SCR-9
FT /note="TRUNCATED"
FT misc_difference 35
FT /note="Gly substituted by Glu (SCR-8)"
FT EP-512733-A.
FT 11-NOV-1992.
FT 28-APR-1992: 303826.
FT 03-MAY-1991: US-695514.
FT (UNIW ) UNIV WASHINGTON.
```

```
PI Atkinson JP, Hourcade D, Kiyoch M;
DR MPI; 92-375009/46.
PT Complement activity regulator protein analogues - useful for
PT treating auto:immune diseases, to suppress transplant rejection,
PT for diagnosis etc.
PS Claim 11; Fig 2 and R11810; 23pp; English.
CC The cDNA clone designated CRI-4 was described in J. Exp. Med. (1988)
CC 168:1255-1270. It encodes the first 8 and a half amino terminal
CC SCRs of CRI. The invention concerns analogues of "regulator of
CC complement activation" proteins or truncated, hybrid or recombinant
CC forms of them. CRI-4 is a preferred truncated form and a number of
CC specified substitution variants of it are claimed. Positions 35 and
CC 37 of SCR-1 and the corresponding positions in SCR-8 have been
CC identified as important in C4b binding. The specification does not
CC contain the CRI-4 sequence; the sequence given here was constructed
CC from the full-length CRI amino acid sequence having GENSESQ
CC accession number R11810 and descriptions in the disclosure.
CC Sequence 543 AA;

SQ
Query Match 95.9%; Score 118; DB 1; Length 543;
Best Local Similarity 95.7%; Pred. No. 6.8e-11;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FELVGEPSIVTSNDQVGIMSG 23
Db 164 FELVGEPSIVCTSNDDQVGIMSG 186

RESULT 7
R28545 ID R28545 standard; peptide: 543 AA.
AC R28545:
DT 19-MAR-1993 (first entry)
DE CRI-4 (37Y) analogue.
KW Short consensus repeat; regulator of complement activation;
KW C3b binding; C4b binding; human complement type 1 receptor.
OS Homo sapiens.
PI Key Location/Qualifiers
FH 1.60
FT /label= SCR-1
FT region 61..122
FT /label= SCR-2
FT region 451..510
FT /label= SCR-8
FT region 511..543
FT /label= SCR-9
FT /note="TRUNCATED"
FT misc_difference 37
FT /note="Ser substituted by Tyr (SCR-8)"
FT EP-512733-A.
FT 11-NOV-1992.
FT 28-APR-1992: 303826.
FT 03-MAY-1991: US-695514.
FT (UNIW ) UNIV WASHINGTON.
FT Atkinson JP, Hourcade D, Kiyoch M;
FT MPI; 92-375009/46.
FT Complement activity regulator protein analogues - useful for
FT treating auto:immune diseases, to suppress transplant rejection,
FT for diagnosis etc.
PS Claim 11; Fig 2 and R11810; 23pp; English.
CC The cDNA clone designated CRI-4 was described in J. Exp. Med. (1988)
CC 168:1255-1270. It encodes the first 8 and a half amino terminal
CC SCRs of CRI. The invention concerns analogues of "regulator of
CC complement activation" proteins or truncated, hybrid or recombinant
CC forms of them. CRI-4 is a preferred truncated form and a number of
CC specified substitution variants of it are claimed. Positions 35 and
CC 37 of SCR-1 and the corresponding positions in SCR-8 have been
CC identified as important in C4b binding. The specification does not
CC contain the CRI-4 sequence; the sequence given here was constructed
CC from the full-length CRI amino acid sequence having GENSESQ
CC accession number R11810 and descriptions in the disclosure.
CC Sequence 543 AA;
```

Query Match 95.9%; Score 118; DB 1; Length 543;  
Best Local Similarity 95.7%; Pred. No. 6.8e-11;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FELVGEPSIYSTSNDQVGIMSG 23  
Db 164 FELVGEPSICTSNDQVGIMSG 186

RESULT 8  
R28546  
ID R28546 standard; peptide; 543 AA.  
AC R28546;  
DT 19-MAR-1993 (first entry)  
DE CRI-4 (44T, 47D, 49L) analogue.  
KW Short consensus repeat; regulator of complement activation;  
C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT region 1..60  
FT /label= SCR-1  
FT region 61..122  
FT /label= SCR-2  
FT region 451..510  
FT /label= SCR-8  
FT region 511..543  
FT /label= SCR-9  
FT /note= "TRUNCATED"  
FT misc\_difference 44  
FT /note= "Ile substituted by Thr (SCR-8)"  
FT misc\_difference 47  
FT /note= "Lys substituted by Asp (SCR-8)"  
FT misc\_difference 49  
FT /note= "Ser substituted by Leu (SCR-8)"  
FN EP-512733-A.  
PN 11-NOV-1992.  
PD 28-APR-1992; 303826.  
PR 03-MAY-1991; US-695514.  
PA (UNIM ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Krych M;  
DR WPI; 92-375009/46.  
PT Complement activity regulator protein analogues - useful for  
treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11; Fig 2 and R11810; 23pp; English.  
CC The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)  
168:1255-1270. It encodes the first 8 and a half amino terminal  
SCRS of CRI. The invention concerns analogues of "regulator of  
complement activation" proteins or truncated, hybrid or recombinant  
forms of them. CRI-4 is a preferred truncated form and a number of  
specified substitution variants of it are claimed in which certain  
positions in SCR-1 which have been identified as important for the  
degree of C3b- and C4b-binding are substituted by amino acids from  
the corresponding positions in SCR-8. The specification does not  
contain the CRI-4 sequence; the sequence given here was constructed  
from the full-length CRI amino acid sequence having GENSEO  
accession number R11810 and descriptions in the disclosure.  
CC Sequence 543 AA;

Query Match 95.9%; Score 118; DB 1; Length 543;  
Best Local Similarity 95.7%; Pred. No. 6.8e-11;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FELVGEPSIYSTSNDQVGIMSG 23  
Db 164 FELVGEPSICTSNDQVGIMSG 186

RESULT 9  
R28547

ID R28547 standard; peptide; 543 AA.  
AC R28547;  
DT 19-MAR-1993 (first entry)  
DE CRI-4 (52S, 53S, 54P) analogue.  
KW Short consensus repeat; regulator of complement activation;  
C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.

FH Key Location/Qualifiers  
FT region 1..60  
FT /label= SCR-1  
FT region 61..122  
FT /label= SCR-2  
FT region 451..510  
FT /label= SCR-8  
FT region 511..543  
FT /label= SCR-9  
FT /note= "TRUNCATED"  
FT misc\_difference 52  
FT /note= "Thr substituted by Ser (SCR-8)"  
FT misc\_difference 53  
FT /note= "Gly substituted by Ser (SCR-8)"  
FT misc\_difference 54  
FT /note= "Ala substituted by Pro (SCR-8)"  
FN EP-512733-A.  
PN 11-NOV-1992.  
PD 28-APR-1992; 303826.  
PR 03-MAY-1991; US-695514.  
PA (UNIM ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Krych M;  
DR WPI; 92-375009/46.  
PT Complement activity regulator protein analogues - useful for  
treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11; Fig 2 and R11810; 23pp; English.  
CC The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)  
168:1255-1270. It encodes the first 8 and a half amino terminal  
SCRS of CRI. The invention concerns analogues of "regulator of  
complement activation" proteins or truncated, hybrid or recombinant  
forms of them. CRI-4 is a preferred truncated form and a number of  
specified substitution variants of it are claimed in which certain  
positions in SCR-1 which have been identified as important for the  
degree of C3b- and C4b-binding are substituted by amino acids from  
the corresponding positions in SCR-8. The specification does not  
contain the CRI-4 sequence; the sequence given here was constructed  
from the full-length CRI amino acid sequence having GENSEO  
accession number R11810 and descriptions in the disclosure.  
CC Sequence 543 AA;

Query Match 95.9%; Score 118; DB 1; Length 543;  
Best Local Similarity 95.7%; Pred. No. 6.8e-11;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FELVGEPSIYSTSNDQVGIMSG 23  
Db 164 FELVGEPSICTSNDQVGIMSG 186

RESULT 10  
R28548  
ID R28548 standard; peptide; 543 AA.  
AC R28548;  
DT 19-MAR-1993 (first entry)  
DE CRI-4 (57V, 59K) analogue.  
KW Short consensus repeat; regulator of complement activation;  
C3b binding; C4b binding; human complement type 1 receptor.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT region 1..60  
FT /label= SCR-1  
FT region 61..122  
FT /label= SCR-2  
FT region 451..510

```
FT      /label= SCR-8
FT      511..543
FT      /label= SCR-9
FT      /note= "RUNCATED"
FT      misc_difference 57
FT      /note= "Arg substituted by Val (SCR-8)"
FT      misc_difference 59
FT      /note= "Arg substituted by Lys (SCR-8)"
PN      EP-512733-A.
PD      11-NOV-1992.
PF      28-APR-1992: 303826.
PR      03-MAY-1991: US-695514.
PA      (UNITV ) UNIV WASHINGTON.
PI      Atkinson JP, Hourcade D, Krych M;
PT      Complement activity regulator protein analogues - useful for
PT      treating auto:immune diseases, to suppress transplant rejection,
PT      for diagnosis etc.
PS      Claim 11; Fig 2 and R1810; 23pp; English.
CC      The CDNA clone designated CRI-4 was described in J.Exp.Med.(1988)
CC      168:1255-1270. It encodes the first 8 and a half amino terminal
CC      SCRs of CRI. The invention concerns analogues of "regulator of
CC      complement activation" proteins or truncated, hybrid or recombinant
CC      forms of them. CRI-4 is a preferred truncated form and a number of
CC      specified substitution variants of it are claimed in which certain
CC      positions in SCR-1 which have been identified as important for the
CC      degree of C3b- and C4b-binding are substituted by amino acids from
CC      the corresponding positions in SCR-8. The specification does not
CC      contain the CRI-4 sequence; the sequence given here was constructed
CC      from the full-length CRI amino acid sequence having GENESEQ
CC      accession number R1810 and descriptions in the disclosure.
SQ      Sequence 543 AA;
```

```
Query Match          95.9%; Score 118; DB 1; Length 543;
Best Local Similarity 95.7%; Pred. No. 6.8e-11;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY      1 FELVGEPSIYTSNDQVGIMSG 23
      |||||
DB      164 FELVGEPSIYCTSNDOVGIMSG 186

RESULT 11
R28549
ID      R28549 standard; peptide: 543 AA.
AC      R28549;
DE      19-MAR-1993 (first entry)
DE      CRI-4 (64k, 65f) analogue.
KW      short consensus repeat; regulator of complement activation;
KW      C3b binding; C4b binding; human complement type 1 receptor.
OS      Homo sapiens.
FH      Key
FT      region
FT      1..60
FT      /label= SCR-1
FT      61..122
FT      /label= SCR-2
FT      451..510
FT      /label= SCR-8
FT      511..543
FT      /label= SCR-9
FT      /note= "TRUNCATED"
FT      misc_difference 64
FT      /note= "Arg substituted by Lys (SCR-9)"
FT      misc_difference 65
FT      /note= "Asn substituted by Thr (SCR-9)"
PN      EP-512733-A.
PD      11-NOV-1992.
PF      28-APR-1992: 303826.
PR      03-MAY-1991: US-695514.
PA      (UNITV ) UNIV WASHINGTON.
PI      Atkinson JP, Hourcade D, Krych M;
PT      Complement activity regulator protein analogues - useful for
PT      treating auto:immune diseases, to suppress transplant rejection,
PT      for diagnosis etc.
PS      Claim 11; Fig 2 and R1810; 23pp; English.
CC      The CDNA clone designated CRI-4 was described in J.Exp.Med.(1988)
CC      168:1255-1270. It encodes the first 8 and a half amino terminal
CC      SCRs of CRI. The invention concerns analogues of "regulator of
CC      complement activation" proteins or truncated, hybrid or recombinant
CC      forms of them. CRI-4 is a preferred truncated form and a number of
CC      specified substitution variants of it are claimed in which certain
CC      positions in SCR-2 which have been identified as important for the
CC      degree of C3b- and C4b-binding are substituted by amino acids from
CC      the corresponding positions in SCR-9. The specification does not
CC      contain the CRI-4 sequence; the sequence given here was constructed
CC      from the full-length CRI amino acid sequence having GENESEQ
CC      accession number R1810 and descriptions in the disclosure.
SQ      Sequence 543 AA;
```

```
PT      Complement activity regulator protein analogues - useful for
PT      treating auto:immune diseases, to suppress transplant rejection,
PT      for diagnosis etc.
PS      Claim 11; Fig 2 and R1810; 23pp; English.
CC      The CDNA clone designated CRI-4 was described in J.Exp.Med.(1988)
CC      168:1255-1270. It encodes the first 8 and a half amino terminal
CC      SCRs of CRI. The invention concerns analogues of "regulator of
CC      complement activation" proteins or truncated, hybrid or recombinant
CC      forms of them. CRI-4 is a preferred truncated form and a number of
CC      specified substitution variants of it are claimed in which certain
CC      positions in SCR-2 which have been identified as important for the
CC      degree of C3b- and C4b-binding are substituted by amino acids from
CC      the corresponding positions in SCR-9. The specification does not
CC      contain the CRI-4 sequence; the sequence given here was constructed
CC      from the full-length CRI amino acid sequence having GENESEQ
CC      accession number R1810 and descriptions in the disclosure.
SQ      Sequence 543 AA;
```

```
Query Match          95.9%; Score 118; DB 1; Length 543;
Best Local Similarity 95.7%; Pred. No. 6.8e-11;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY      1 FELVGEPSIYTSNDQVGIMSG 23
      |||||
DB      164 FELVGEPSIYCTSNDOVGIMSG 186

RESULT 12
R28550
ID      R28550 standard; peptide: 543 AA.
AC      R28550;
DE      19-MAR-1993 (first entry)
DE      CRI-4 (64k) analogue.
KW      short consensus repeat; regulator of complement activation;
KW      C3b binding; C4b binding; human complement type 1 receptor.
OS      Homo sapiens.
FH      Key
FT      region
FT      1..60
FT      /label= SCR-1
FT      61..122
FT      /label= SCR-2
FT      451..510
FT      /label= SCR-8
FT      511..543
FT      /label= SCR-9
FT      /note= "TRUNCATED"
FT      misc_difference 64
FT      /note= "Arg substituted by Lys (SCR-9)"
PN      EP-512733-A.
PD      11-NOV-1992.
PF      28-APR-1992: 303826.
PR      03-MAY-1991: US-695514.
PA      (UNITV ) UNIV WASHINGTON.
PI      Atkinson JP, Hourcade D, Krych M;
PT      Complement activity regulator protein analogues - useful for
PT      treating auto:immune diseases, to suppress transplant rejection,
PT      for diagnosis etc.
PS      Claim 11; Fig 2 and R1810; 23pp; English.
CC      The CDNA clone designated CRI-4 was described in J.Exp.Med.(1988)
CC      168:1255-1270. It encodes the first 8 and a half amino terminal
CC      SCRs of CRI. The invention concerns analogues of "regulator of
CC      complement activation" proteins or truncated, hybrid or recombinant
CC      forms of them. CRI-4 is a preferred truncated form and a number of
CC      specified substitution variants of it are claimed in which certain
CC      positions in SCR-2 which have been identified as important for the
CC      degree of C3b- and C4b-binding are substituted by amino acids from
CC      the corresponding positions in SCR-9. The specification does not
CC      contain the CRI-4 sequence; the sequence given here was constructed
CC      from the full-length CRI amino acid sequence having GENESEQ
CC      accession number R1810 and descriptions in the disclosure.
SQ      Sequence 543 AA;
```

Query Match 95.9%; Score 118; DB 1; Length 543;  
Best Local Similarity 95.7%; Pred. No. 6.8e-11;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FELVGEPSIYTSNDQVGIMSG 23  
|||||  
Db 164 FELVGEPSIYCTSNDDQVGIMSG 186

RESULT 13  
R28551  
ID R28551 standard; peptide; 543 AA.  
AC R28551;  
DT 19-MAR-1993 (first entry)  
DE CRI-4 (65T) analogue.  
KW Short consensus repeat; regulator of complement activation;  
OS Homo sapiens.

FH Key Location/Qualifiers  
FT region 1..60  
FT /label= SCR-1  
FT region 61..122  
FT /label= SCR-2  
FT region 451..510  
FT /label= SCR-8  
FT region 511..543  
FT /label= SCR-9  
FT /note= "TRUNCATED"

FT misc-difference 64  
FT /note= "Arg substituted by Lys (SCR-9)"

FT misc-difference 65  
FT /note= "Asn substituted by Thr (SCR-9)"

PN EP-512733-A.  
PD 11-NOV-1992.  
PF 28-APR-1992; 303826.  
PR 03-MAY-1991; US-695514.  
PA (UNITW ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Krych M;  
DR WPI: 92-375009/46.  
PT Complement activity regulator protein analogues - useful for  
PT treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11; Fig 2 and R11810; 23pp; English.  
CC The cDNA clone designated CRI-4 was described in J. Exp. Med. (1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCRs of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-2 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-9. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQ  
CC accession number R11810 and descriptions in the disclosure.  
SQ Sequence 543 AA;

Query Match 95.9%; Score 118; DB 1; Length 543;  
Best Local Similarity 95.7%; Pred. No. 6.8e-11;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FELVGEPSIYTSNDQVGIMSG 23  
|||||  
Db 164 FELVGEPSIYCTSNDDQVGIMSG 186

RESULT 14  
R28552  
ID R28552 standard; peptide; 543 AA.  
AC R28552;

DT 19-MAR-1993 (first entry)  
DE CRI-4 (78T, 79D) analogue.  
KW Short consensus repeat; regulator of complement activation;  
OS Homo sapiens.

FH Key Location/Qualifiers  
FT region 1..60  
FT /label= SCR-1  
FT region 61..122  
FT /label= SCR-2  
FT region 451..510  
FT /label= SCR-8  
FT region 511..543  
FT /label= SCR-9  
FT /note= "TRUNCATED"

FT misc-difference 78  
FT /note= "Lys substituted by Thr (SCR-9)"

FT misc-difference 79  
FT /note= "Gly substituted by Asp (SCR-9)"

PN EP-512733-A.  
PD 11-NOV-1992.  
PF 28-APR-1992; 303826.  
PR 03-MAY-1991; US-695514.  
PA (UNITW ) UNIV WASHINGTON.  
PI Atkinson JP, Hourcade D, Krych M;  
DR WPI: 92-375009/46.  
PT Complement activity regulator protein analogues - useful for  
PT treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11; Fig 2 and R11810; 23pp; English.  
CC The cDNA clone designated CRI-4 was described in J. Exp. Med. (1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCRs of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-2 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-9. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQ  
CC accession number R11810 and descriptions in the disclosure.  
SQ Sequence 543 AA;

Query Match 95.9%; Score 118; DB 1; Length 543;  
Best Local Similarity 95.7%; Pred. No. 6.8e-11;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FELVGEPSIYTSNDQVGIMSG 23  
|||||  
Db 164 FELVGEPSIYCTSNDDQVGIMSG 186

RESULT 15  
R28553  
ID R28553 standard; peptide; 543 AA.  
AC R28553;  
DT 19-MAR-1993 (first entry)  
DE CRI-4 (85R, 87N) analogue.  
KW Short consensus repeat; regulator of complement activation;  
OS Homo sapiens.

FH Key Location/Qualifiers  
FT region 1..60  
FT /label= SCR-1  
FT region 61..122  
FT /label= SCR-2  
FT region 451..510  
FT /label= SCR-8  
FT region 511..543  
FT /label= SCR-9  
FT /note= "TRUNCATED"

FT misc\_difference 85 /note= "Gln substituted by Arg (SCR-9)"  
 FT misc\_difference 87 /note= "Lys substituted by Asn (SCR-9)"  
 FT  
 PN EP-512733-A.  
 PD 11-NOV-1992.  
 PR 28-APR-1992; 303826.  
 PR 03-MAY-1991; US-695514.  
 PA (UNIV ) UNIV WASHINGTON.  
 PI ALKINSON JP, Hourcade D, Krych M;  
 DR WPI: 92-375009/46.  
 PT Complement activity regulator protein analogues - useful for  
 PT treating auto-immune diseases, to suppress transplant rejection,  
 PT for diagnosis etc.  
 PS Claim 11; Fig 2 and R11810; 23pp; English.  
 CC The cDNA clone designated CRI-4 was described in J. Exp. Med. (1988)  
 CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
 CC SCRs of CRI. The invention concerns analogues of "regulator of  
 CC complement activation" proteins or truncated, hybrid or recombinant  
 CC forms of them. CRI-4 is a preferred truncated form and a number of  
 CC specified substitution variants of it are claimed in which certain  
 CC positions in SCR-2 which have been identified as important for the  
 CC degree of C3b- and C4b-binding are substituted by amino acids from  
 CC the corresponding positions in SCR-9. The specification does not  
 CC contain the CRI-4 sequence; the sequence given here was constructed  
 CC from the full-length CRI amino acid sequence having GENBSEQ  
 CC accession number R11810 and descriptions in the disclosure.  
 SQ Sequence 543 AA;

Query Match 95.9%; Score 118; DB 1; Length 543;  
 Best Local Similarity 95.7%; Pred. No. 6.8e-11;

Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FELVGEPSITSTNDQVGIMSG 23  
 |||

Db 164 FELVGEPSITCTSNDDQVGIMSG 186

Search completed: July 18, 1999, 00:42:55  
 Job time: 8588 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 06:07:57 ; Search time 39.54 Seconds  
(without alignments)  
5.740 Million cell updates/sec

Title: US-09-142-043-8

Perfect score: 123

Sequence: 1 FELVGEPSIYTSNDQVIGMSG 23

Scoring table: BLOSUM62

Searched: 106577 seqs, 9868381 residues

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/PCTUS9\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/Backfilest1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	118	95.9	197	2	US-08-356-361-27
2	118	95.9	76	2	US-08-356-361-28
3	118	95.9	254	2	US-08-356-361-29
4	118	95.9	254	2	US-08-356-361-30
5	118	95.9	133	2	US-08-356-361-31
6	118	95.9	197	2	US-08-769-967A-27
7	118	95.9	76	2	US-08-769-967A-28
8	118	95.9	254	2	US-08-769-967A-29
9	118	95.9	254	2	US-08-769-967A-30
10	118	95.9	133	2	US-08-769-967A-31
11	64.5	52.4	324	1	US-08-310-416A-14
12	64.5	52.4	324	2	US-08-888-171-14
13	64.5	52.4	323	2	US-08-435-149-2
14	64.5	52.4	577	2	US-08-435-149-3
15	48	39.0	340	4	5256642-2
16	48	39.0	340	4	5472939-2
17	46	37.4	310	4	5256642-10
18	46	37.4	310	4	5472939-10
19	45	36.6	446	1	US-07-781-254A-2
20	45	36.6	446	1	US-07-781-254A-3
21	45	36.6	444	1	US-07-626-618A-20
22	45	36.6	444	1	US-08-333-977-20
23	45	36.6	10	4	5256642-18
24	45	36.6	10	4	5472939-18
25	44	35.8	254	1	US-08-310-416A-13
26	44	35.8	293	1	US-08-310-416A-16
27	44	35.8	169	1	US-08-310-416A-18
28	44	35.8	140	2	US-08-664-596B-2
29	44	35.8	140	2	US-08-738-367-2
30	44	35.8	240	2	US-08-738-367-7
31	44	35.8	377	2	US-08-528-057-2
32	44	35.8	370	2	US-08-528-057-42
33	44	35.8	324	2	US-08-528-057-44
34	44	35.8	324	2	US-08-528-057-46
35	44	35.8	254	2	US-08-888-171-13
36	44	35.8	293	2	US-08-888-171-16
37	44	35.8	169	2	US-08-888-171-18
38	44	35.8	254	2	US-08-435-149-1
39	41	33.3	888	1	US-08-445-640-35

40	41	33.3	1664	2	US-08-642-846-2	Sequence 2, Appl1
41	41	33.3	260	4	5256642-5	Patent No. 5256642
42	41	33.3	250	4	5256642-6	Patent No. 5256642
43	41	33.3	260	4	5472939-5	Patent No. 5472939
44	41	33.3	250	4	5472939-6	Patent No. 5472939
45	40	32.5	606	2	US-08-883-534-3	Sequence 3, Appl1

## ALIGNMENTS

```
RESULT 1
US-08-356-361-27
; Sequence 27, Application US/08356361
; Patent No. 5833989
; GENERAL INFORMATION:
; APPLICANT: Smith, Richard A.G.
; APPLICANT: Dodd, Ian
; APPLICANT: Freeman Mary A.
; APPLICANT: Mossakowska, Danuta E.I.
; TITLE OF INVENTION: No. 5833989e1 Compounds
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham-Corporate Intellectual Property
; STREET: P.O. Box 1539
; CITY: King of Prussia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/356,361
; FILING DATE: 03-Jul-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jervais, Herbert H.
; REGISTRATION NUMBER: 31,171
; REFERENCE/DOCKET NUMBER: P30423
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (610) 270-5019
; TELEFAX: (610) 270-5090
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 197 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: N-terminal
US-08-356-361-27

Query Match          95.9%; Score 118; DB 2; Length 197;
Best Local Similarity 95.7%; Pred. No. 3.8e-11;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FELVGEPSIYTSNDQVIGMSG 23
DB 165 FELVGEPSIYTSNDQVIGMSG 187

RESULT 2
US-08-356-361-28
; Sequence 28, Application US/08356361
; Patent No. 5833989
; GENERAL INFORMATION:
; APPLICANT: Smith, Richard A.G.
; APPLICANT: Dodd, Ian
; APPLICANT: Freeman Mary A.
; APPLICANT: Mossakowska, Danuta E.I.
```

TITLE OF INVENTION: No. 5833989el Compounds  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: SmithKline Beecham-Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,361  
FILING DATE: 03-Jul-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jervis, Herbert H.  
REGISTRATION NUMBER: 31,171  
REFERENCE/DOCKET NUMBER: P30423  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5019  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 76 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-356-361-28

Query Match 95.9%; Score 118; DB 2; Length 76;  
Best Local Similarity 95.7%; Pred. No. 1.3e-11;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FELVGEPSIYTSNDQVGIMSG 23  
|||||||  
Db 44 FELVGEPSIYCTSDNDQVGIMSG 66

RESULT 3  
US-08-356-361-29  
Sequence 29, Application US/08356361  
Patent No. 5833989  
GENERAL INFORMATION:  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman Mary A.  
APPLICANT: Mossakowska, Danuta E.I.  
TITLE OF INVENTION: No. 5833989el Compounds  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: SmithKline Beecham-Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,361  
FILING DATE: 03-Jul-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jervis, Herbert H.

REGISTRATION NUMBER: 31,171  
REFERENCE/DOCKET NUMBER: P30423  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5019  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-356-361-29

Query Match 95.9%; Score 118; DB 2; Length 254;  
Best Local Similarity 95.7%; Pred. No. 5.1e-11;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FELVGEPSIYTSNDQVGIMSG 23  
|||||||  
Db 165 FELVGEPSIYCTSDNDQVGIMSG 187

RESULT 4  
US-08-356-361-30  
Sequence 30, Application US/08356361  
Patent No. 5833989  
GENERAL INFORMATION:  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman Mary A.  
APPLICANT: Mossakowska, Danuta E.I.  
TITLE OF INVENTION: No. 5833989el Compounds  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: SmithKline Beecham-Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,361  
FILING DATE: 03-Jul-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jervis, Herbert H.  
REGISTRATION NUMBER: 31,171  
REFERENCE/DOCKET NUMBER: P30423  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5019  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-356-361-30

Query Match 95.9%; Score 118; DB 2; Length 254;  
Best Local Similarity 95.7%; Pred. No. 5.1e-11;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FELVGEPSIYTSNDQVGIMSG 23



DB 165 FELVGEPSICTSNDQVGIMSG 187

## RESULT 5

US-08-356-361-31  
Sequence 31, Application US/08356361  
Patent No. 5833989  
GENERAL INFORMATION:  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman Mary A.  
APPLICANT: Mossakowska, Danuta E.I.  
TITLE OF INVENTION: No. 5833989el Compounds  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham-Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,361  
FILING DATE: 03-Jul-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jervis, Herbert H.  
REGISTRATION NUMBER: 31,171  
REFERENCE/DOCKET NUMBER: P30423  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5019  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ. ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 133 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-356-361-31

Query Match 95.9%; Score 118; DB 2; Length 133;  
Best local Similarity 95.7%; Pred. No. 2.4e-11;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FELVGEPSICTSNDQVGIMSG 23  
DB 44 FELVGEPSICTSNDQVGIMSG 66

## RESULT 6

US-08-769-967A-27  
Sequence 27, Application US/08769967A  
Patent No. 5859223  
GENERAL INFORMATION:  
APPLICANT: Mossakowska, Danuta E.I.  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman, Anne Mary  
TITLE OF INVENTION: Soluble CRI Derivatives  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania

COUNTRY: USA  
ZIP: 19406

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,967A  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,569  
FILING DATE: 15-May-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: P30423C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5364  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ. ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 197 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-769-967A-27

Query Match 95.9%; Score 118; DB 2; Length 197;  
Best local Similarity 95.7%; Pred. No. 3.8e-11;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FELVGEPSICTSNDQVGIMSG 23  
DB 165 FELVGEPSICTSNDQVGIMSG 187

## RESULT 7

US-08-769-967A-28  
Sequence 28, Application US/08769967A  
Patent No. 5859223  
GENERAL INFORMATION:  
APPLICANT: Mossakowska, Danuta E.I.  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman, Anne Mary  
TITLE OF INVENTION: Soluble CRI Derivatives  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,967A  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,569  
FILING DATE: 15-May-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954

REFERENCE/DOCKET NUMBER: P30423C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5364  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 76 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-769-967A-28

Query Match 95.9%; Score 118; DB 2; Length 76;  
Best Local Similarity 95.7%; Pred. No. 1,3e-11;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FELVGEPSIYSTSNDQVGIMSG 23  
|||||  
DB 44 FELVGEPSICTSNDQVGIMSG 66

RESULT 8  
US-08-769-967A-29  
Sequence 29, Application US/08769967A  
Patent No. 5859223  
GENERAL INFORMATION:  
APPLICANT: Mossakowska, Danuta E.I.  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman, Anne Mary  
TITLE OF INVENTION: Soluble CRI Derivatives  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Smithline Beecham Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,967A  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,569  
FILING DATE: 15-May-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: P30423C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5364  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-769-967A-29

Query Match 95.9%; Score 118; DB 2; Length 254;  
Best Local Similarity 95.7%; Pred. No. 5.1e-11;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FELVGEPSIYSTSNDQVGIMSG 23  
|||||  
DB 165 FELVGEPSICTSNDQVGIMSG 187

RESULT 9  
US-08-769-967A-30  
Sequence 30, Application US/08769967A  
Patent No. 5859223  
GENERAL INFORMATION:  
APPLICANT: Mossakowska, Danuta E.I.  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman, Anne Mary  
TITLE OF INVENTION: Soluble CRI Derivatives  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Smithline Beecham Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,967A  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,569  
FILING DATE: 15-May-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: P30423C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5364  
TELEFAX: (610) 270-5090  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-769-967A-30

Query Match 95.9%; Score 118; DB 2; Length 254;  
Best Local Similarity 95.7%; Pred. No. 5.1e-11;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FELVGEPSIYSTSNDQVGIMSG 23  
|||||  
DB 165 FELVGEPSICTSNDQVGIMSG 187

RESULT 10  
US-08-769-967A-31  
Sequence 31, Application US/08769967A  
Patent No. 5859223  
GENERAL INFORMATION:  
APPLICANT: Mossakowska, Danuta E.I.  
APPLICANT: Smith, Richard A.G.  
APPLICANT: Dodd, Ian  
APPLICANT: Freeman, Anne Mary  
TITLE OF INVENTION: Soluble CRI Derivatives  
NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporate Intellectual Property  
STREET: P.O. Box 1539  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,967A  
FILING DATE:  
CLASSIFICATION: 536

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,569  
FILING DATE: 15-May-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: P30423C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5364  
TELEFAX: (610) 270-5090

INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 133 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal

US-08-769-967A-31

Query Match	95.9%	Score 118;	DB 2;	Length 133;
Best Local Similarity	95.7%	Pred. No. 2.4e-11;		
Matches 22;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;

RESULT 11  
 US-08-310-416A-14  
 Sequence 14, Application US/08310416A  
 Patent No. 5679546  
 GENERAL INFORMATION:  
 APPLICANT: Jone-Long Ko et al.  
 TITLE OF INVENTION: CHIMERIC PROTEINS WHICH BLOCK  
 TITLE OF INVENTION: COMPLEMENT ACTIVATION  
 NUMBER OF SEQUENCES: 19  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: Fish & Richardson P.C.  
 STREET: 225 Franklin Street  
 CITY: Boston  
 STATE: Massachusetts  
 COUNTRY: U.S.A.  
 ZIP: 02110-2804  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
 COMPUTER: IBM PS/2 Model 502 or 515SX  
 OPERATING SYSTEM: MS-DOS (Version 5.0)  
 SOFTWARE: Wordperfect (Version 5.1)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/310.416A  
 FILING DATE: 22-SEP-1994  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Paul T. Clark  
 REGISTRATION NUMBER: 30,162

```

: REFERENCE/DOCKET NUMBER: 06180/0050001
:
: TELECOMMUNICATION INFORMATION:
:
: TELEPHONE: (617) 542-5070
:
: TELEFAX: (617) 542-8906
:
: TELEX: 200154
:
: INFORMATION FOR SEO ID NO: 14:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 324 amino acids
:
: TYPE: amino acid
:
: STRANDEDNESS: not relevant
:
: TOPOLOGY: linear
:
: MOLECULE TYPE: protein
:
: US-08-310-416A-14

```

Query Match	52.4%;	Score 64.5;	DB 1;	Length 324;
Best Local Similarity	56.5%;	Pred. NO. 0.011;		
Matches 13; Conservative	4;	Mismatches 5;	Indels 1;	Gaps 1;

```

RESULT 12
US-08-888-171-14
: Sequence 14, Application US/08088171
: Patent No. 5851528
:
: GENERAL INFORMATION:
: APPLICANT: Jone-Long, Ko
: APPLICANT: Higgins, Paul J.
: APPLICANT: Yeh, C. Grace
: TITLE OF INVENTION: METHODS OF INHIBITING COMPLEMENT
: TITLE OF INVENTION: ACTIVATION
: NUMBER OF SEQUENCES: 19
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson, P.C.
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: MA
:
: COUNTRY: US
: ZIP: 02110-2804
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
:
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows95
: SOFTWARE: FastSEO for Windows Version 2.0
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/888,171
: FILING DATE: 03-JUL-1997
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/310,416
: FILING DATE: 22-SEP-1994
:
: ATTORNEY/AGENT INFORMATION:
: NAME: Freeman, John W.
: REGISTRATION NUMBER: 29,066
: REFERENCE/DOCKET NUMBER: 06180/005002
:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/542-507
: TELEFAX: 617/542-890
:
: TELEX: 200154
:
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 324 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
:
: MOLECULE TYPE: protein
:
: US-08-888-171-14

```

Query Match	52.48;	Score 64.5;	DB 2;	Length 324;
Best Local Similarity	56.58;	Pred. No. 0.011;		
Matches 13; Conservative	4;	Mismatches 5;	Indels 1;	Gaps 1;



Search completed: July 18, 1999, 06:07:58  
job time: 300 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 06:07:06 ; Search time 44.47 Seconds

(without alignments)  
19.375 Million cell updates/sec

Title: US-09-142-043-8

Perfect score: 123  
Sequence: 1 FELVGEPSIYSTSNDQVGIMSG 23

Scoring table: BLOSUM62

Searched: 116738 seqs, 37460341 residues

Database : PIR\_58:\*\*

1: pir1:\*\*  
2: pir2:\*\*  
3: pir3:\*\*  
4: pir4:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	118	95.9	2014	2	I36936	complement recepto
2	118	95.9	661	2	I36937	complement recepto
3	118	95.9	2489	2	I73012	complement C3b/C4b
4	118	95.9	2039	2	A28507	complement C3b/C4b
5	112	91.1	482	2	A34924	complement C3b/C4b
6	99	80.5	433	2	A30550	complement C3b/C4b
7	99	80.5	440	2	A43519	complement recepto
8	91	74.0	497	2	JC2054	complement regulat
9	87	70.7	676	2	A45900	complement C3d rec
10	71	57.7	1091	1	PL0009	complement C3d/Eps
11	66.5	54.1	340	2	I56234	decay-accelerating
12	64.5	52.4	381	1	B26359	decay-accelerating
13	64.5	52.4	440	2	A26359	decay-accelerating
14	62	50.4	1025	2	A43526	complement C3d/Eps
15	61	49.6	330	2	I55975	X/Y protein - mous
16	56	45.5	597	2	S53711	C4b alpha chain p
17	51	41.5	583	2	S56680	beta-fructofuranos
18	51	41.5	634	2	A55683	I-iduronidase (EC
19	50	40.7	582	2	S37047	beta-fructofuranos
20	50	40.7	580	2	S49508	beta-fructofuranos
21	50	40.7	311	2	C70125	Lambda CII stablil
22	49	39.8	597	1	NBHUC4	C4b-binding protei
23	49	39.8	587	2	S36231	beta-fructofuranos
24	49	39.8	87	2	S20912	regulatory protein
25	49	39.8	87	2	S29614	whib protein - Str
26	48	39.0	87	2	S33906	beta-fructofuranos
27	47	38.2	592	2	S56681	beta-fructofuranos
28	47	38.2	592	2	S61503	C4b protein alpha
29	47	38.2	558	2	S57953	probable membrane
30	46.5	37.8	219	2	S61188	dopamine receptor
31	45	36.6	446	1	DYK1D3	cadherin 3 proteor
32	45	36.6	829	1	IJHUCP	cytochrome P450 1A
33	45	36.6	521	2	S34184	dopamine receptor
34	45	36.6	446	2	I48322	NADH dehydrogenase
35	45	36.6	217	2	S28754	hypothetical prote
36	44.5	36.2	206	2	C70950	C4b-binding protei
37	44	35.8	469	1	NBMS4	probable whib2 pro
38	44	35.8	89	2	C70847	calcium-binding pr
39	44	35.8	313	2	G69498	

40 44 35.8 369 2 I57998 membrane cofactor  
41 44 35.8 384 2 S01896 membrane cofactor  
42 44 35.8 377 2 I54479 membrane cofactor  
43 44 35.8 349 2 G02913 sperm CD46 - human  
44 43 35.4 494 2 A47059 sucrose Scrb - Sta  
45 43 35.0 369 1 ESBYPC 3',5'-cyclic-nucle

## ALIGNMENTS

RESULT 1  
I36936  
complement receptor 1 - chimpanzee (fragment)  
C:Species: Pan troglodytes (chimpanzee)  
C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 07-Aug-1998  
C/Accession: I36936  
R:Birmingham, D.J.; Shen, X.P.; Hourcade, D.; Nickells, M.W.; Atkinson, J.P.  
J. Immunol. 153, 691-700, 1994  
A>Title: Primary sequence of an alternatively spliced form of CRI. Candidate for the  
A:Reference number: I36935; MUID:94292799  
A:Accession: I36936  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-2014 <RES>  
A:Cross-references: GB:I24920; NID:9551564; PID:9557725  
C:Superfamily: complement factor H repeat homology  
F:18-74/Domain: complement factor H repeat homology <FH01>  
F:79-136/Domain: complement factor H repeat homology <FH02>  
F:141-207/Domain: complement factor H repeat homology <FH03>  
F:333-391/Domain: complement factor H repeat homology <FH12>  
F:1041-1107/Domain: complement factor H repeat homology <FH26>  
F:1749-1815/Domain: complement factor H repeat homology <FH27>

Query Match 95.9%; Score 118; DB 2; Length 2014;  
Best Local Similarity 95.7%; Pred. No. 6.6e-10;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FELVGEPSIYSTSNDQVGIMSG 23

DB 180 FELVGEPSIYCTSNDQVGIMSG 202

RESULT 2  
I36937  
complement receptor 1 - chimpanzee (fragment)  
C:Species: Pan troglodytes (chimpanzee)  
C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 07-Aug-1998  
C/Accession: I36937  
R:Birmingham, D.J.; Shen, X.P.; Hourcade, D.; Nickells, M.W.; Atkinson, J.P.  
J. Immunol. 153, 691-700, 1994  
A>Title: Primary sequence of an alternatively spliced form of CRI. Candidate for the  
A:Reference number: I36935; MUID:94292799  
A:Accession: I36937  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-661 <RES>  
A:Cross-references: GB:I24921; NID:9557726; PID:9557727  
C:Superfamily: complement factor H repeat homology  
F:18-74/Domain: complement factor H repeat homology <FH01>  
F:79-136/Domain: complement factor H repeat homology <FH02>  
F:141-207/Domain: complement factor H repeat homology <FH03>  
F:213-268/Domain: complement factor H repeat homology <FH04>  
F:333-391/Domain: complement factor H repeat homology <FH22>  
F:396-462/Domain: complement factor H repeat homology <FH27>

Query Match 95.9%; Score 118; DB 2; Length 661;  
Best Local Similarity 95.7%; Pred. No. 1.8e-10;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FELVGEPSIYSTSNDQVGIMSG 23

```

Db      180 FELGEPSTICTSNDQVIGMSG 202

RESULT  3
173012 Complement C3b/C4b receptor (allotype S) precursor - human
N:Alternate names: complement receptor type 1 (CRI); surface glycoprotein CD35
C:Species: Homo sapiens (man)
C:Date: 02-Jul-1996 #sequence revision 02-Jul-1996 #text_change 24-Sep-1998
C:Accession: 173012; A47602; S03291
R:VLK, D.P.; Wong, W.W.
J. Immunol. 151, 6214-6224, 1993
A:Title: Structure of the gene for the F allele of complement receptor type 1 and sequen
A:Reference number: 156203; MUID:94065175
A:Accession: 173012
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2489 <RES>
A:Cross-references: GB:117418; NID:g306678; PID:g451303
R:Wong, W.W.; Cahill, J.M.; Rosen, M.D.; Kennedy, C.A.; Bonaccio, E.T.; Morris, M.J.; W
J. Exp. Med. 169, 847-863, 1989
A:Title: Structure of the human CRI gene. Molecular basis of the structural and quantita
A:Reference number: A47602
A:Accession: A47602
A:Molecule type: DNA
A:Residues: 1-41 <MON>
A:Cross-references: GB:X14893
R:Hourcade, D.; Mesner, D.R.; Atkinson, J.P.; Holers, V.M.
J. Exp. Med. 168, 1235-1270, 1988
A:Title: Identification of an alternative polyadenylation site in the human C3b/C4b rece
Type 1.
A:Reference number: S03291; MUID:89010527
A:Accession: S03291
A:Molecule type: mRNA
A:Residues: 26-584 <HOU>
A:Cross-references: EMBL:X14362; NID:g30197; PID:g736240
A:Experimental source: clone CRI-4
C:Genetics:
A:Gene: GDB:CR1; CD35
A:Cross-references: GDB:119800; OMIM:120620
A:Map position: 1q32-1q32
A:Introns: 41/1; 101/1; 134/2; 163/1; 296/1; 355/1; 389/2; 418/1; 491/1; 551/1; 584/2; 6
; 1484/2; 1513/1; 1646/1; 1705/1; 1739/2; 1768/1; 1844/1; 1904/1; 1937/2; 1966/1; 2099/1
C:Keywords: duplication; glycoprotein; transmembrane protein
F:1-41/Domain: signal sequence #status predicted <SIG>
F:42-584/Product: complement C3b/C4b receptor, secreted #status predicted <MAT>
F:43-99/Domain: complement factor H repeat homology <FH01>
F:104-161/Domain: complement factor H repeat homology <FH02>
F:166-232/Domain: complement factor H repeat homology <FH03>
F:238-293/Domain: complement factor H repeat homology <FH04>
F:297-353/Domain: complement factor H repeat homology <FH05>
F:356-416/Domain: complement factor H repeat homology <FH06>
F:421-487/Domain: complement factor H repeat homology <FH07>
F:554-611/Domain: complement factor H repeat homology <FH08>
F:616-682/Domain: complement factor H repeat homology <FH09>
F:808-866/Domain: complement factor H repeat homology <FHXB>
F:1004-1061/Domain: complement factor H repeat homology <FH09>
F:1066-1132/Domain: complement factor H repeat homology <FH10>
F:1138-1193/Domain: complement factor H repeat homology <FH11>
F:1197-1253/Domain: complement factor H repeat homology <FH12>
F:1258-1316/Domain: complement factor H repeat homology <FH13>
F:1321-1387/Domain: complement factor H repeat homology <FH14>
F:1393-1449/Domain: complement factor H repeat homology <FH15>
F:1454-1511/Domain: complement factor H repeat homology <FH16>
F:1516-1582/Domain: complement factor H repeat homology <FH17>
F:1588-1643/Domain: complement factor H repeat homology <FH18>
F:1647-1703/Domain: complement factor H repeat homology <FH19>
F:1708-1766/Domain: complement factor H repeat homology <FH20>
F:1771-1837/Domain: complement factor H repeat homology <FH21>
F:1846-1902/Domain: complement factor H repeat homology <FH22>
F:1907-1964/Domain: complement factor H repeat homology <FH23>

```

```

F:1969-2035/Domain: complement factor H repeat homology <FHXC>
F:2100-2156/Domain: complement factor H repeat homology <FH24>
F:2161-2219/Domain: complement factor H repeat homology <FHXD>

Query Match      95.98; Score 118; DB 2; Length 2489;
Best Local Similarity 95.7%; Pred. No. 8,4e-10;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Cy      1 FELGEPSTICTSNDQVIGMSG 23
Db      205 FELGEPSTICTSNDQVIGMSG 227

RESULT  4
A28507 Complement C3b/C4b receptor precursor, membrane-bound (allotype F) - human
N:Alternate names: complement receptor type 1 (CRI); surface glycoprotein CD35
C:Species: Homo sapiens (man)
C:Date: 19-Nov-1988 #sequence revision 06-Sep-1996 #text_change 10-Sep-1997
C:Accession: S03843; A28507; 156203; A24748; B24748; C24748
R:Klickstein, L.B.; Barlow, T.J.; Miletic, V.; Rabson, L.D.; Smith, J.A.; Fearon, D.T
J. Exp. Med. 168, 1699-1717, 1988
A:Title: Identification of distinct C3b and C4b recognition sites in the human C3b/C4
A:Reference number: S03843; MUID:89035992
A:Accession: S03843
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-2039 <KL1>
A:Cross-references: EMBL:Y00816; NID:g30185; PID:g30186
R:Klickstein, L.B.; Wong, W.W.; Smith, J.A.; Weis, J.H.; Wilson, J.G.; Fearon, D.T.
J. Exp. Med. 165, 1095-1112, 1987
A:Title: Human C3b/C4b receptor (CRI). Demonstration of long homologous repeating dom
A:Reference number: A28507; MUID:87168191
A:Accession: A28507
A:Molecule type: mRNA
A:Residues: 503-771, 'FV', 774-2039 <KL2>
R:VLK, D.P.; Wong, W.W.
J. Immunol. 151, 6214-6224, 1993
A:Title: Structure of the gene for the F allele of complement receptor type 1 and seq
A:Reference number: 156203; MUID:94065175
A:Accession: 156203
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-683, 'X', 685-1021, 'V', 1023-1614, 'V', 1616-1826, 'R', 1828-1849, 'D', 1851-187
R:Wong, W.W.; Klickstein, L.B.; Smith, J.A.; Weis, J.H.; Fearon, D.T.
Proc. Natl. Acad. Sci. U.S.A. 82, 7711-7715, 1985
A:Title: Identification of a partial cDNA clone for the human receptor for complement
A:Reference number: A94073; MUID:86067975
A:Accession: A24748
A:Molecule type: mRNA
A:Residues: 311-333; 729-745; 831-845 <WON>
C:Genetics:
A:Gene: GDB:CR1; CD35
A:Cross-references: GDB:119800; OMIM:120620
A:Map position: 1q32-1q32
A:Introns: 41/1; 101/1; 134/2; 163/1; 296/1; 355/1; 389/2; 418/1; 491/1; 551/1; 584/2
; 1487/2; 1516/1; 1649/1; 1708/1; 1742/2; 1771/1; 1847/1; 1906/1; 1968/1; 1976/1; 200
C:Keywords: duplication; glycoprotein; transmembrane protein
F:1-41/Domain: signal sequence #status predicted <SIG>
F:42-2039/Product: complement C3b/C4b receptor, membrane-bound #status predicted <MAT>
F:43-99/Domain: complement factor H repeat homology <FH01>
F:104-161/Domain: complement factor H repeat homology <FH02>
F:166-232/Domain: complement factor H repeat homology <FH03>
F:238-293/Domain: complement factor H repeat homology <FH04>
F:297-353/Domain: complement factor H repeat homology <FH05>
F:358-416/Domain: complement factor H repeat homology <FH06>
F:421-487/Domain: complement factor H repeat homology <FH07>
F:493-549/Domain: complement factor H repeat homology <FH07>
F:554-611/Domain: complement factor H repeat homology <FH08>
F:616-682/Domain: complement factor H repeat homology <FH09>

```



F:688-743/Domain: complement factor H repeat homology <FH10>  
F:747-803/Domain: complement factor H repeat homology <FH11>  
F:808-866/Domain: complement factor H repeat homology <FH12>  
F:871-937/Domain: complement factor H repeat homology <FH13>  
F:943-999/Domain: complement factor H repeat homology <FH14>  
F:1004-1061/Domain: complement factor H repeat homology <FH15>  
F:1066-1132/Domain: complement factor H repeat homology <FH16>  
F:1138-1193/Domain: complement factor H repeat homology <FH17>  
F:1197-1253/Domain: complement factor H repeat homology <FH18>  
F:1258-1316/Domain: complement factor H repeat homology <FH19>  
F:1321-1387/Domain: complement factor H repeat homology <FH20>  
F:1396-1452/Domain: complement factor H repeat homology <FH21>  
F:1457-1514/Domain: complement factor H repeat homology <FH22>  
F:1519-1585/Domain: complement factor H repeat homology <FH23>  
F:1591-1646/Domain: complement factor H repeat homology <FH24>  
F:1650-1706/Domain: complement factor H repeat homology <FH25>  
F:1711-1769/Domain: complement factor H repeat homology <FH26>  
F:1774-1840/Domain: complement factor H repeat homology <FH27>  
F:1848-1904/Domain: complement factor H repeat homology <FH28>  
F:1909-1965/Domain: complement factor H repeat homology <FH29>

Query Match 95.9%; Score 118; DB 2; Length 2039;  
Best Local Similarity 95.7%; Pred. No. 6.7e-10;

Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FELVGPSTSTNDPOVGWISG 23  
|||||

DB 205 FELVGPSTSTNDPOVGWISG 227

RESULT 5  
A34924  
complement C3b/C4b receptor-like protein precursor - human

C:Species: Homo sapiens (man)

C:Date: 27-Jul-1990 #sequence\_revision 11-Apr-1997 #text\_change 11-Apr-1997

C:Accession: A34924; S03292

R:Hourcade, D.; Miesner, D.R.; Bee, C.; Zeldes, W.; Atkinson, J.P.

J. Biol. Chem. 265, 974-980, 1990

A:Title: Duplication and divergence of the amino-terminal coding region of the complement

A:Reference number: A34924; MUID:90110163

A:Accession: A34924

A:Molecule type: DNA

A:Residues: 1-479 <H0U>

A:Cross-references: GB:J05195

R:Hourcade, D.; Miesner, D.R.; Atkinson, J.P.; Holiers, V.M.

J. Exp. Med. 168, 1255-1270, 1988

A:Title: Identification of an alternative polyadenylation site in the human C3b/C4b receptor

A:Reference number: S03291; MUID:89010527

A:Accession: S03292

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: FFFFAFR, 452-482 <H02>

A:Cross-references: EMBL:X14360

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1988

C:Superfamily: complement factor H repeat homology

F:43-99/Domain: complement factor H repeat homology <FH01>

F:104-161/Domain: complement factor H repeat homology <FH02>

F:166-332/Domain: complement factor H repeat homology <FH03>

F:238-293/Domain: complement factor H repeat homology <FH04>

F:297-353/Domain: complement factor H repeat homology <FH05>

F:358-416/Domain: complement factor H repeat homology <FH06>

F:421-478/Domain: complement factor H repeat homology <FH07>

Query Match 91.1%; Score 112; DB 2; Length 482;  
Best Local Similarity 91.3%; Pred. No. 1e-09;

Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FELVGPSTSTNDPOVGWISG 23  
|||||

DB 205 FELVGPSTSTNDPOVGWISG 227

RESULT 6  
A30550  
complement C3b/C4b receptor precursor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 03-Mar-1989 #sequence\_revision 03-Mar-1989 #text\_change 12-May-1995

C:Accession: A30550

R:Paul, M.S.; Aegeerter, M.; O'Brien, S.E.; Kurtz, C.B.; Weis, J.H.

J. Immunol. 142, 582-589, 1989

A:Title: The murine complement receptor gene family. Analysis of mCR1 gene products a

A:Reference number: A30550; MUID:89093944

A:Accession: A30550

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-433 <PAU>

C:Superfamily: complement factor H repeat homology

F:42-98/Domain: complement factor H repeat homology <FH1>

F:103-160/Domain: complement factor H repeat homology <FH2>

F:165-231/Domain: complement factor H repeat homology <FH3>

F:237-293/Domain: complement factor H repeat homology <FH4>

F:299-355/Domain: complement factor H repeat homology <FH5>

Query Match 80.5%; Score 99; DB 2; Length 433;  
Best Local Similarity 69.6%; Pred. No. 9.5e-08;

Matches 16; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 FELVGPSTSTNDPOVGWISG 23  
|||||

DB 204 FELVGPSTSTNDPOVGWISG 226

RESULT 7  
A43519  
complement receptor CR1 precursor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 28-Oct-1992 #sequence\_revision 30-Jan-1993 #text\_change 12-May-1995

C:Accession: A43519

R:Paul, M.S.; Aegeerter, M.; Cepek, K.; Miller, M.D.; Weis, J.H.

J. Immunol. 144, 1988-1996, 1990

A:Title: The murine complement receptor gene family. The genomic and transcriptional

A:Reference number: A43519; MUID:90171600

A:Accession: A43519

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-440 <PAU>

A:Cross-references: GB:M34164

A:Note: the authors translated the codon GGC for residue 21 as Ala, and CAG for resid

C:Superfamily: complement factor H repeat homology

F:42-98/Domain: complement factor H repeat homology <FH1>

F:103-160/Domain: complement factor H repeat homology <FH2>

F:165-231/Domain: complement factor H repeat homology <FH3>

F:237-293/Domain: complement factor H repeat homology <FH4>

F:299-355/Domain: complement factor H repeat homology <FH5>

Query Match 80.5%; Score 99; DB 2; Length 440;  
Best Local Similarity 69.6%; Pred. No. 9.7e-08;

Matches 16; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 FELVGPSTSTNDPOVGWISG 23  
|||||

DB 204 FELVGPSTSTNDPOVGWISG 226

RESULT 8

JC2054  
complement regulatory protein, 512 antigen precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 27-Jun-1994 #sequence\_revision 27-Jun-1994 #text\_change 22-Nov-1996

C:Accession: JC2054; PC2027

R:Sakurada, C.; Seno, H.; Dohl, N.; Takizawa, H.; Nonaka, M.; Okada, N.; Gada, H.

	Query Match	74.0%	Score 91;	DB 2;	Length 497;
	Best Local Similarity		Pred. No.	1.9e-06;	
	Matches 17; Conservative	2;	Mismatches	4;	Indels
			Gaps	0;	
OY	1 FELVGEPSITSTNSNDQYGIMSG	23			
DG	200 FNLVGEPSIHCTSIDQGVGWMSG	222			

	Query Match	70.7%;	Score 87;	DB 2;	Length 676;
	Best Local Similarity	72.7%;	Pred. No.	1.2e-05;	
	Matches 16; Conservative	3;	Mismatches	3;	Indels 0; Gaps 0;
OY	1 FETVGEPISTGTSNDPOVGITSS	22			
	::   ::   ::    :				
Dd	184 FDVLGEKSIYTCTSKDNOVGIAW	205			

RESULT 10

PL0009

Complement C3b/Epstein-Barr virus receptor precursor - human

N:Alternate names: complement receptor 2; CR2/CD21

C:Species: Homo sapiens (man)

C:Date: 30-Jun-1992 #sequence, revision 07-Jul-1995 #text, change 05-Sep-1997

C:Accession: J10028; A33958; A32036; A24319; B24319; C24319; D24319; E24319; F24319; R:Wels, J.J.; Toothaker, L.E.; Smith, J.A.; Wels, J.H.; Fearon, D.T.

J. Exp. Med. 167, 1047-1066, 1988

A:Title: Structure of the human B lymphocyte receptor for C3d and the Epstein-Barr virus

A:Reference number: J10028; MUID:88171282

A:Accession: J10028

A:Molecule type: mRNA

A:Residues: 1-1091 <ME1>

A>Note: nucleotides 1566-1625 are missing from Figure 1; therefore, residues 522-542

R:Moore, M.D.; Cooper, N.R.; Tack, B.F.; Nemerow, G.R.

Proc. Natl. Acad. Sci. U.S.A. 84, 9194-9198, 1987

A:Title: Molecular cloning of the CDNA encoding the Epstein-Barr virus/C3d receptor (

A:Reference number: A39958; MUID:88097454

A:Accession: A39958

A:Molecule type: mRNA

A:Residues: 1-456, 'G', 457-658, 'R', 646-669, 'R', 671-816, 'NCSAEYLKAMILERAF', 835-840, 'L'

A:Cross-references: GB:J03565; NID:9181919; PID:9181920

R:Fujisaku, A.; Harley, J.B.; Frack, M.B.; Gruner, B.A.; Frazier, B.; Holers, V.M.

J. Biol. Chem. 264, 2118-2125, 1989

A:Title: Genomic organization and polymorphisms of the human C3d/Epstein-Barr virus r

A:Reference number: A32036; MUID:89123277

A:Accession: A32036

A:Molecule type: mRNA

A:Residues: 1-456, 'G', 457-658, 718-1050, 'I', 1052-1060, 'E', 1062-1091 <FUT>

A:Cross-references: GB:J04463

R:Wels, J.J.; Fearon, D.T.; Klickstein, L.B.; Wong, W.W.; Richards, S.A.; De Bruyn KC

Proc. Natl. Acad. Sci. U.S.A. 83, 5639-5643, 1986

A:Title: Identification of a partial cDNA clone for the C3d/Epstein-Barr virus recept

Of complement.

A:Reference number: A94114; MUID:86287311

A:Accession: A24319

A:Molecule type: Protein

A:Residues: 226-230, 'XILQ', 257-267, 332-341, 583-591, 'Q', 593, 'D', 595-596, 728-735 <ME2>

A:Experimental source: B-Lymphoblastoid cell lines SB and Raj1

C:Genetics:

A:Gene: GDB:CR2

A:Cross-references: GDB:119802; OMIM:120650

A:Map position: 1q32-1q32

C:Superfamily: complement C3d/Epstein-Barr virus receptor; complement factor H repeat

C:Keywords: alternative splicing; duplication; glycoprotein; transmembrane protein

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-1091/Product: complement receptor 2 (16-repeat form) #status predicted <MA1>

F:21-658, 718-1091/Product: complement receptor 2 (15-repeat form) #status predicted <

F:91-146/Domain: complement factor H repeat homology <FH01>

F:91-146/Domain: complement factor H repeat homology <FH02>

F:154-210/Domain: complement factor H repeat homology <FH03>

F:215-771/Domain: complement factor H repeat homology <FH04>

F:276-342/Domain: complement factor H repeat homology <FH05>

F:351-406/Domain: complement factor H repeat homology <FH06>

F:410-465/Domain: complement factor H repeat homology <FH07>

F:470-521/Domain: complement factor H repeat homology <FH08>

F:526-592/Domain: complement factor H repeat homology <FH09>

F:601-656/Domain: complement factor H repeat homology <FH10>

F:660-716/Domain: complement factor H repeat homology <FH11>

F:720-772/Domain: complement factor H repeat homology <FH12>

F:777-837/Domain: complement factor H repeat homology <FH13>

F:846-901/Domain: complement factor H repeat homology <FH14>

F:909-965/Domain: complement factor H repeat homology <FH1.5>  
F:970-1026/Domain: complement factor H repeat homology <FH1.6>  
F:1034-1056/Domain: transmembrane #status predicted <TM>  
F:1057-1091/Domain: intracellular #status predicted <INT>  
F:121,127,294,372,622,698,858,881,919/Binding site: carbohydrate (Asn) (covalent) #status predicted <B>

Query Match	57.7%	Score 71:	DB 1:	Length 1091:
Best Local Similarity	56.5%	Pred:	No. 0.0062:	
Matches 13, Conservative		3, Mismatches	7, Indels	0: Gaps 0:

```
Qy 1 FELVGEPSIYSTNDQVGIMSG 23
    | | : | | : | | |
Db 565 FSLIGESTIRCTSDNQERGTWSG 587
```

```

RESULT 11
156234
decay:accelerating factor - orangutan (fragment)

```

C.Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 07-Aug-1998  
C.Accession: I56234  
R.Nickells, M.W.; Alvarez, J.I.; Lublin, D.M.; Atkinson, J.P.

A:Title: Characterization of DAF-2, a high molecular weight form of decay-accelerating factor (DAF) in human neutrophils  
A:Reference number: 156234  
A:Accession: T8324

Query Match	54.1%	Score	66.5	DB	2	Length	340
Best Local Similarity	56.5%	Pred. No.	0.0078				
Matches	13	Conservative	4	Mismatches	5	Indels	1
						Gaps	1

```
QY      1 FELVGEPSIYSTSNDDQVGIMSG 23
      : ::|| ||| | |||: | |||
Db     216 YTMIGEHSIYCTVNDDE-GEWSG 23
```

RESULT 12

decay-accelerating factor (GPI-anchored splice form) precursor - human  
N:Alternate names: CD55; DAF splice form 2; decay-accelerating factor membrane-bound form  
C:Species: Homo sapiens (man)  
C:Date: 05-Oct-1998 #sequence\_revision 16-Aug-1996 #text\_change 24-Apr-1998

R:Geras, I.W., Davlitz, M.A.; Rhee, L.; Weddell, G.; Martin Jr., D.W.; Nussenzweig, V.  
Nature 325, 545-549, 1987  
A>Title: Cloning of decay-accelerating factor suggests novel use of splicing to generate  
A;Reference number: A26359; MUID:87115845

A:Molecule type: mRNA  
A:Residues: 5-79, 'T', 81-84, 'M', 86-381 <MED>  
A:Cross-references: GB:M15799; NID:g181462; PID:g181463  
J:Moran, P., Raab, H., Kohl, W., Carras, I. M.  
J. Biol. Chem. 266, 1250-1257, 1991  
A:Title: Glycophospholipid membrane anchor attachment. Molecular analysis of the cleavage  
J:Reference number: A39101; MUID:91093238

A:Accession: A39101  
A:Molecule type: Protein  
A:Residues: 338-352 <MOR>  
R:Ruthlin, D.M.; Mallinson, G.; Poole, J.; Reid, M.E.; Thompson, E.S.; Ferdman, B.R.; Blood 84, 1276-1282, 1989  
A:Title: Molecular basis of reduced or absent expression of decay-accelerating factor  
A:Reference number: 152594

Query Match	52.4%	Score 64.5;	DB 1;	length 381;
Best Local Similarity	56.5%;	Pred. No. 0.018;		
Matches 13; Conservative	4;	Mismatches 5;	Indels 1;	Gaps 1;

```
QY      1 FELVGEPSIYSTSNDQVGIWSG 23
        |::||| | |::| |||
Db      257 FTMIGEHSIYCTVNND-GEWSG 278
```

## RESULT

decay-accelerating factor splice form 1 precursor - human  
N:Alternate names: decay-accelerating factor 5; decay-accelerating factor secreted for C:Species: Homo sapiens (hmn)  
C:Date: 05-Oct-1988 #sequence\_revision 05-Oct-1988 #text\_change 29-Aug-1997  
C:Accession: A26359, A39702, S16187, S21338, A27258  
R:Caras, I.W., Davitz, M.A., Rhee, L., Weddell, G., Martin Jr., D.W., Nussenzweig, V. Nature 325, 545-549, 1987  
A:Title: Cloning of decay-accelerating factor suggests novel use of splicing to generate  
Reference number: A26359; MUID:87115845

A:Cross-references: GB:M30142  
R:Ewulonu, U.R.; Ravi, L.; Medof, M.E.  
Proc. Natl. Acad. Sci. U.S.A. 88, 4675-4679, 1991  
A:Title: Characterization of the decay-accelerating factor gene promoter region  
Reference number: A39702; MJD:91271256

A:Accession: A39702  
 A:Molecule type: DNA  
 A:Residues: 1-79, 'T', 81-104 <EMU>  
 A:Cross-references: GB:M64356  
 A:Note: the authors translated the codon AGC for residue 85 as Met  
 R:Nakano, Y.; Sugita, Y.; Ishikawa, Y.; Choi, N.H.; Tobe, T.; Tomita, M.  
 Biochim. Biophys. Acta 1074, 326-330, 1991  
 A:Title: Isolation of two forms of decay-accelerating factor (DAF) from human urine.  
 A:Reference number: S16187; MUID:91291869  
 A:Accession: S16187  
 A>Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 35-47 <BIO>  
 R:Nakano, Y.; Sumida, K.; Kikuta, N.; Mura, N.H.; Tobe, T.; Tomita, M.  
 Biochim. Biophys. Acta 1116, 235-240, 1992  
 A:Title: Complete determination of disulfide bonds localized within the short consensus  
 A:Reference number: S23138  
 A:Accession: S23138  
 A>Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 35-41; 65-68; 79-81; 93-103; 128-134; 143-145; 155-159; 162-168; 188-192; 203-204; 211  
 R:Sugita, Y.; Negro, T.; Matsuda, T.; Sakamoto, T.; Tomita, M.  
 J. Biochem. 100, 143-150, 1985  
 A:Title: Improved method for the isolation and preliminary characterization of human DAF  
 A:Reference number: A27258; MUID:87008461  
 A:Accession: A27258  
 A:Molecule type: protein  
 A:Residues: 35, 'X', 37, 'G', 39-51, 'P', 53-55, 'X', 57-58, 'X', 60-63 <SUG>  
 A:Note: Gly-35 and Leu-38 were also found  
 C:Genetics:  
 A:Gene: GDB:DAF  
 A:Cross-references: GDB:119088; OMIM:125240  
 A:Map position: 1932-1932  
 C:Superfamily: decay-accelerating factor; complement factor H repeat homology  
 C:Keywords: alternative splicing; glycoprotein  
 F:1-34/Domain: signal sequence #status predicted <SIG>  
 F:35-440/Product: decay-accelerating factor 1 #status predicted <SIG>  
 F:36-94/Domain: complement factor H repeat homology <FH01>  
 F:98-158/Domain: complement factor H repeat homology <FH02>  
 F:163-220/Domain: complement factor H repeat homology <FH03>  
 F:225-283/Domain: complement factor H repeat homology <FH04>  
 F:95/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 52.4%; Score 64.5; DB 2; Length 440;  
 Best Local Similarity 56.5%; Pred. No. 0.022;  
 Matches 13; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

OY 1 FELVGEPSIYSTNDQVIGMSG 23  
 DB 257 FTMIGHSIYCTVNDDE-GEVMSG 278

# RESULT 14

A43526  
 Complement C3d/Epstein-Barr virus receptor 2 precursor - mouse  
 N:Alternate names: complement receptor type 2  
 C:Species: Mus musculus (house mouse)  
 C:Date: 28-Oct-1992 #sequence revision 28-Oct-1992 #text\_change 31-Oct-1997  
 C:Accession: A43526; A35358; A32215; A45802; B32215  
 R:Fingerioth, J.D.  
 J. Immunol. 144, 3458-3467, 1990  
 A:Title: Comparative structure and evolution of murine CR2. The homolog of the human C3d  
 A:Reference number: A43526; MUID:90229735  
 A:Accession: A43526  
 A:Molecule type: mRNA  
 A:Residues: 1-1025 <FIN>  
 A:Cross-references: GB:M35684; EMBL:J04153; NID:g192687; PID:g192688  
 R:Molina, H.; Kinoshita, T.; Inoue, K.; Carel, J.C.; Holers, V.M.  
 J. Immunol. 145, 2974-2983, 1990  
 A:Title: A molecular and immunochemical characterization of mouse CR2. Evidence for a si  
 A:Reference number: A43538; MUID:91010789  
 A:Accession: A43538

A:Molecule type: mRNA  
 A:Residues: 12-305, 'T', 307-519, 'A', 521-1025 <MOL>  
 A:Cross-references: GB:M61132; NID:g192692; PID:g192693  
 R:Fingerioth, J.D.; Benedict, M.A.; Levy, D.N.; Strominger, J.L.  
 Proc. Natl. Acad. Sci. U.S.A. 86, 242-246, 1989  
 A:Title: Identification of murine complement receptor type 2.  
 A:Reference number: A32215; MUID:89098890  
 A:Accession: A32215  
 A:Molecule type: mRNA  
 A:Residues: 343-401; 991-1025 <F12>  
 A:Cross-references: GB:J04153  
 R:Kurtz, C.B.; Paul, M.S.; Aegeerter, M.; Weis, J.J.; Weis, J.H.  
 J. Immunol. 143, 2058-2067, 1989  
 A:Title: Murine complement receptor gene family. Identification and characterization  
 A:Reference number: A45802  
 A:Accession: A45802  
 A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 'E', 100-101, 292-961, 964-1025 <KUR>  
 A:Cross-references: GB:M29281; NID:g192685; PID:g387131  
 A:Note: the authors failed to translate GGA for residue 421 as Gly, and CCA for resid  
 A:Note: the authors translated the codon CAC for residue 727 as Asn  
 C:Superfamily: complement C3d/Epstein-Barr virus receptor; complement factor H repeat  
 C:Keywords: alternative splicing; duplication; glycoprotein; receptor; transmembrane  
 F:1-11/Domain: signal sequence #status predicted <SIG>  
 F:12-1025/Product: complement C3d/Epstein-Barr virus receptor 2 (15-repeat form) #sta  
 F:12-973/Domain: extracellular #status predicted <EXT>  
 F:14-73/Domain: complement factor H repeat homology <FH02>  
 F:82-138/Domain: complement factor H repeat homology <FH03>  
 F:146-202/Domain: complement factor H repeat homology <FH04>  
 F:207-263/Domain: complement factor H repeat homology <FH05>  
 F:268-334/Domain: complement factor H repeat homology <FH06>  
 F:343-398/Domain: complement factor H repeat homology <FH07>  
 F:402-458/Domain: complement factor H repeat homology <FH08>  
 F:463-514/Domain: complement factor H repeat homology <FH09>  
 F:519-585/Domain: complement factor H repeat homology <FH10>  
 F:594-649/Domain: complement factor H repeat homology <FH11>  
 F:654-704/Domain: complement factor H repeat homology <FH12>  
 F:709-769/Domain: complement factor H repeat homology <FH13>  
 F:778-833/Domain: complement factor H repeat homology <FH14>  
 F:841-897/Domain: complement factor H repeat homology <FH15>  
 F:902-958/Domain: complement factor H repeat homology <FH16>  
 F:968-989/Domain: transmembrane #status predicted <TM>  
 F:990-1025/Domain: intracellular #status predicted <INT>

Query Match 50.4%; Score 62; DB 2; Length 1025;  
 Best Local Similarity 47.8%; Pred. No. 0.14;  
 Matches 11; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

OY 1 FELVGEPSIYSTNDQVIGMSG 23  
 DB 307 FTLIGKINTCTGSKTGMSG 329

# RESULT 15

I55975  
 X/Y protein - mouse (fragment)  
 C:Species: Mus musculus (house mouse)  
 C:Date: 26-Jul-1996 #sequence revision 26-Jul-1996 #text\_change 10-Oct-1997  
 C:Accession: I55975  
 R:Aegerter-Shaw, M.; Cole, J.L.; Klickestein, L.B.; Wong, W.W.; Fearon, D.T.; Talley,  
 J. Immunol. 138, 3488-3494, 1987  
 A:Title: Expansion of the complement receptor gene family: Identification in the mous  
 A:Reference number: I55975; MUID:87196575  
 A:Accession: I55975  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-330 <RES>  
 A:Cross-references: GB:M61179; NID:g202427; PID:g202428  
 C:Superfamily: complement factor H repeat homology  
 F:36-92/Domain: complement factor H repeat homology <FH4>  
 F:188-244/Domain: complement factor H repeat homology <FH1>

F:249-306/Domain: complement factor H repeat homology <FH02>

Query Match 49.68; Score 61; DB 2; Length 330;  
Best Local Similarity 64.3%; Pred. No. 0.054;  
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 10 YSTSNDDQVGIVSG 23  
| | | | : | | | |  
Db 12 YCTSNDDGEIGVWSG 25

Search completed: July 18, 1999, 06:07:07  
Job time: 253 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 11:37:53 ; Search time 31.53 Seconds

(Without alignments)  
19.579 Million cell updates/sec

Title: US-09-142-043-8

Perfect score: 123

Sequence: 1 FELVGEPSRSTSTNDQ , JMSG 23

Scoring table: BLOSUM62

Searched: 74019 seqs, 26840295 residues

Database: SwissProt\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	118	95.9	2039	1	CRI_HUMAN
2	71	57.7	1033	1	CR2_HUMAN
3	66.5	54.1	340	1	DAF_PONPY
4	64.5	52.4	381	1	DAF_HUMAN
5	62	50.4	1025	1	CR2_MOUSE
6	55	44.7	390	1	DAF1_MOUSE
7	52	42.3	590	1	INVA_MAIZE
8	51	41.5	634	1	IDUA_MOUSE
9	51	41.5	583	1	INV3_DAUCA
10	49	39.8	597	1	C4BP_HUMAN
11	47	38.2	558	1	C4BP_RAT
12	47	38.2	407	1	DAF2_MOUSE
13	47	38.2	592	1	INW2_DAUCA
14	45	36.6	829	1	CAD3_HUMAN
15	45	36.6	446	1	D3DR_MOUSE
16	45	36.6	446	1	D3DR_RAT
17	45	36.6	217	1	NU2M_MYTED
18	44	35.8	469	1	C4BP_MOUSE
19	44	35.8	917	1	PAN_HUMAN
20	44	35.8	377	1	MCP_HUMAN
21	43.5	35.4	494	1	SCRB_STAXY
22	43	35.0	610	1	C4BP_BOVIN
23	43	35.0	651	1	CORO_YEAST
24	43	35.0	534	1	DBH2_YEAST
25	43	35.0	534	1	PSB2_SYNT3
26	43	35.0	595	1	YR85_SCHPO
27	43	35.0	548	1	PR2N_RENSA
28	42.5	34.6	548	1	AMT4_PSESA
29	42	34.1	551	1	AMT4_PSESA
30	42	34.1	548	1	AMT4_PSESA
31	42	34.1	350	1	CCMH_ECOLI
32	42	34.1	920	1	PAN_MOUSE
33	42	34.1	861	1	FLGB_SOLTU
34	42	34.1	947	1	HLXA_PASSP
35	42	34.1	134	1	SPAK_SALPY
36	42	34.1	612	1	TRPD_DICDI
37	42	34.1	535	1	YB3_YEAST
38	42	34.1	1032	1	YB3_YEAST
39	41.5	33.7	545	1	DEGM_CAEEL
40	41	33.3	400	1	D3DR_CERAE
41	41	33.3	400	1	D3DR_HUMAN
42	41	33.3	958	1	HIG_DROME
43	41	33.3	1664	1	INT1_CANAL

## ALIGNMENTS

RESULT	1	STANDARD:	PRT:	2039 AA.
ID	CRI_HUMAN			
AC	P17927			
DT	01-NOV-1990 (REL. 16, CREATED)			
DT	01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)			
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)			
DE	COMPLEMENT RECEPTOR TYPE 1 PRECURSOR (C3B/C4B RECEPTOR) (CD35			
DE	ANTIGEN).			
GN	CRI OR C3B.			
OS	HOMO SAPIENS (HUMAN).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;			
OC	EUHIERIA; PRIMATES.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 89035992.			
RA	KLICKESTEIN L.B., BARTOW T.J., MILETIC V., RABSON L.D., SMITH J.A.,			
RA	FEARON D.T.;			
RL	J. EXP. MED. 168:1699-1717(1988).			
RN	[2]			
RP	SEQUENCE OF 503-2039 FROM N.A.			
RX	MEDLINE: 87168191.			
RA	KLICKESTEIN L.B., WONG W.W., SMITH J.A., WEIS J.H., WILSON J.G.,			
RA	FEARON D.T.;			
RL	J. EXP. MED. 165:1095-1112(1987).			
RN	[3]			
RP	SEQUENCE OF 761-783; 831-845 AND 1179-1195 FROM N.A.			
RX	MEDLINE: 86067975.			
RA	WONG W.W., KLICKSTEIN L.B., SMITH J.A., WEIS J.H., FEARON D.T.;			
RA	PROC. NATL. ACAD. SCI. U.S.A. 82:7711-7715(1985).			
RL	- FUNCTION: CRI. PRESENT ON ERYTHROCYTES, LEUCOCYTES, GLOMERULAR			
CC	PODCYTES, AND SPLENIC FOLLICULAR DENDRITIC CELLS, MEDIATES THE			
CC	BINDING BY THESE CELLS OF PARTICLES AND IMMUNE COMPLEXES THAT HAVE			
CC	ACTIVATED COMPLEMENT.			
CC	- SUBUNIT: MONOMER.			
CC	- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.			
CC	- POLYMORPHISM: CRI IS RESPONSIBLE FOR THE KNOX BLOOD GROUP SYSTEM.			
CC	- SEVEN SHORT CONSENSUS REPEATS (SCR) CONSTITUTE A LONG HOMOLOGOUS			
CC	REPEAT (LHR). THE N-TERMINAL TWO SCRS OF LHR-A CONTAINED A SITE			
CC	DETERMINING C4 SPECIFICITY, AND THE N-TERMINAL TWO SCRS OF LHR-B			
CC	AND -C EACH HAD A SITE DETERMINING C3 SPECIFICITY.			
CC	- THIS IS THE SEQUENCE OF THE F ALLOTYPIC OF CRI.			
CC	- SIMILARITY: CONTAINS 30 SUSHI (SCR) REPEATS.			
DR	EMBL: M11569; G180995; -			
DR	EMBL: M11617; G180996; -			
DR	EMBL: Y00816; G180997; -			
DR	EMBL: X05309; G8809019; -			
DR	PIR: A28507; A28507.			
DR	PIR: A24748; A24748.			
DR	PIR: B24748; B24748.			
DR	PIR: C24748; C24748.			
DR	PIR: S03843; S03843.			
DR	HSSP: P08603; 1HFI.			
DR	MIN: 10620; -			
KW	COMPLEMENT PATHWAY; GLYCOPROTEIN; TRANSMEMBRANE; REPEAT; SIGNAL;			
KW	RECEPTOR; SUSHI; BLOOD GROUP ANTIGEN.			
FT	SIGNAL	1	41	
FT	CHAIN	42	2039	
FT	DOMAIN	42	1971	
FT	DOMAIN	1972	1996	
FT	DOMAIN	1997	2039	
FT	MOD_RES	42	42	
FT	DOMAIN	42	488	
FT	REPEAT	42	100	
FT	REPEAT A.			
FT	SUSHI AL.			

P35953 Oryctolagi  
O43814 plium sat.

```

FT REPEAT 103 162 SUSHI A2.
FT REPEAT 165 223 SUSHI A3.
FT REPEAT 237 284 SUSHI A4.
FT REPEAT 296 354 SUSHI A5.
FT REPEAT 357 417 SUSHI A6.
FT REPEAT 420 488 SUSHI A7.
FT DOMAIN 492 938 7 x SUSHI (SCR) REPEATS, LONG HOMOLOGOUS
FT REPEAT 492 550 REPEAT B.
FT REPEAT 553 612 SUSHI B1.
FT REPEAT 615 683 SUSHI B2.
FT REPEAT 687 744 SUSHI B3.
FT REPEAT 746 804 SUSHI B4.
FT REPEAT 807 867 SUSHI B5.
FT REPEAT 870 938 SUSHI B6.
FT DOMAIN 942 1388 7 x SUSHI (SCR) REPEATS, LONG HOMOLOGOUS
FT REPEAT 942 1000 REPEAT C.
FT REPEAT 1003 1062 SUSHI C1.
FT REPEAT 1065 1133 SUSHI C2.
FT REPEAT 1137 1194 SUSHI C3.
FT REPEAT 1196 1254 SUSHI C4.
FT REPEAT 1257 1317 SUSHI C5.
FT REPEAT 1320 1388 SUSHI C6.
FT DOMAIN 1395 1846 7 x SUSHI (SCR) REPEATS, LONG HOMOLOGOUS
FT REPEAT 1395 1453 REPEAT D.
FT REPEAT 1456 1515 SUSHI D1.
FT REPEAT 1518 1586 SUSHI D2.
FT REPEAT 1647 1707 SUSHI D3.
FT REPEAT 1710 1770 SUSHI D4.
FT REPEAT 1773 1841 SUSHI D5.
FT DOMAIN 1847 1905 2 x SUSHI (SCR) REPEATS.
FT REPEAT 1908 1966 SUSHI D6.
FT DISULFID 43 86 SUSHI D7.
FT DISULFID 73 99 SUSHI 1.
FT DISULFID 104 145 SUSHI 2.
FT DISULFID 131 161 BY SIMILARITY.
FT DISULFID 166 215 BY SIMILARITY.
FT DISULFID 232 280 BY SIMILARITY.
FT DISULFID 280 293 BY SIMILARITY.
FT DISULFID 297 340 BY SIMILARITY.
FT DISULFID 326 353 BY SIMILARITY.
FT DISULFID 358 400 BY SIMILARITY.
FT DISULFID 386 416 BY SIMILARITY.
FT DISULFID 421 470 BY SIMILARITY.
FT DISULFID 450 487 BY SIMILARITY.
FT DISULFID 493 536 BY SIMILARITY.
FT DISULFID 523 549 BY SIMILARITY.
FT DISULFID 554 595 BY SIMILARITY.
FT DISULFID 581 611 BY SIMILARITY.
FT DISULFID 616 665 BY SIMILARITY.
FT DISULFID 645 682 BY SIMILARITY.
FT DISULFID 688 730 BY SIMILARITY.
FT DISULFID 716 743 BY SIMILARITY.
FT DISULFID 747 790 BY SIMILARITY.
FT DISULFID 776 803 BY SIMILARITY.
FT DISULFID 808 850 BY SIMILARITY.
FT DISULFID 836 866 BY SIMILARITY.
FT DISULFID 871 920 BY SIMILARITY.
FT DISULFID 900 937 BY SIMILARITY.
FT DISULFID 943 986 BY SIMILARITY.
FT DISULFID 973 999 BY SIMILARITY.
FT DISULFID 1004 1045 BY SIMILARITY.
FT DISULFID 1031 1061 BY SIMILARITY.
FT DISULFID 1066 1115 BY SIMILARITY.
FT DISULFID 1095 1132 BY SIMILARITY.
FT DISULFID 1138 1180 BY SIMILARITY.
FT DISULFID 1166 1193 BY SIMILARITY.
FT DISULFID 1197 1240 BY SIMILARITY.

```

```

FT DISULFID 1226 1253 BY SIMILARITY.
FT DISULFID 1258 1300 BY SIMILARITY.
FT DISULFID 1286 1316 BY SIMILARITY.
FT DISULFID 1321 1370 BY SIMILARITY.
FT DISULFID 1350 1387 BY SIMILARITY.
FT DISULFID 1396 1439 BY SIMILARITY.
FT DISULFID 1426 1452 BY SIMILARITY.
FT DISULFID 1457 1498 BY SIMILARITY.
FT DISULFID 1484 1514 BY SIMILARITY.
FT DISULFID 1519 1568 BY SIMILARITY.
FT DISULFID 1548 1585 BY SIMILARITY.
FT DISULFID 1591 1633 BY SIMILARITY.
FT DISULFID 1619 1646 BY SIMILARITY.
FT DISULFID 1650 1693 BY SIMILARITY.
FT DISULFID 1679 1706 BY SIMILARITY.
FT DISULFID 1711 1753 BY SIMILARITY.
FT DISULFID 1739 1769 BY SIMILARITY.
FT DISULFID 1774 1823 BY SIMILARITY.
FT DISULFID 1803 1840 BY SIMILARITY.
FT DISULFID 1848 1891 BY SIMILARITY.
FT DISULFID 1877 1904 BY SIMILARITY.
FT DISULFID 1809 1952 BY SIMILARITY.
FT DISULFID 1938 1965 BY SIMILARITY.
FT CARBOHYD 56 56 POTENTIAL.
FT CARBOHYD 252 252 POTENTIAL.
FT CARBOHYD 410 410 POTENTIAL.
FT CARBOHYD 447 447 POTENTIAL.
FT CARBOHYD 509 509 POTENTIAL.
FT CARBOHYD 578 578 POTENTIAL.
FT CARBOHYD 702 702 POTENTIAL.
FT CARBOHYD 860 860 POTENTIAL.
FT CARBOHYD 897 897 POTENTIAL.
FT CARBOHYD 959 959 POTENTIAL.
FT CARBOHYD 1028 1028 POTENTIAL.
FT CARBOHYD 1110 1152 POTENTIAL.
FT CARBOHYD 1481 1481 POTENTIAL.
FT CARBOHYD 1504 1504 POTENTIAL.
FT CARBOHYD 1534 1534 POTENTIAL.
FT CARBOHYD 1540 1540 POTENTIAL.
FT CARBOHYD 1605 1605 POTENTIAL.
FT CARBOHYD 1763 1763 POTENTIAL.
FT CARBOHYD 1908 1908 POTENTIAL.
SQ SEQUENCE 2039 AA; 223589 MW; 666F9033 CRC32;

Query Match 95.9%; Score 118; DB 1; Length 2039;
Best local Similarity 95.7%; Pred. No. 5.5e-10;
Matches 22; Conservative 0; Mismatches 1; Indels 0;

QY 1 FELVGEPSITSTNDQVGIWSG 23
DB 205 FELVGEPSITCTSDNDQVGIWSG 227

RESULT 2
ID CR2_HUMAN STANDARD; PRT; 1033 AA.
AC P20023;
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE COMPLEMENT RECEPTOR TYPE 2 PRECURSOR (CR2) (COMPLEMENT C3D REC
DE (EPSTEIN-BARR VIRUS RECEPTOR) (CD21 ANTIGEN).
GN CR2 OR C3DR.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89123277.
RA FUJISAKI A., HARLEY J.B., FRANK M.B., GRUNER B.A., FRAZIER B.,
RA HOLERS V.M.;

```





FT DISULFID 778 820 BY SIMILARITY.  
 FT DISULFID 806 833 BY SIMILARITY.  
 FT DISULFID 841 884 BY SIMILARITY.  
 FT DISULFID 902 897 BY SIMILARITY.  
 FT DISULFID 920 945 BY SIMILARITY.  
 FT DISULFID 931 958 BY SIMILARITY.  
 FT CARBOHYD 777 777 POTENTIAL.  
 FT CARBOHYD 113 113 POTENTIAL.  
 FT CARBOHYD 276 276 POTENTIAL.  
 FT CARBOHYD 316 316 POTENTIAL.  
 FT CARBOHYD 364 364 POTENTIAL.  
 FT CARBOHYD 380 380 POTENTIAL.  
 FT CARBOHYD 484 484 POTENTIAL.  
 FT CARBOHYD 527 527 POTENTIAL.  
 FT CARBOHYD 615 615 POTENTIAL.  
 FT CARBOHYD 639 639 POTENTIAL.  
 FT CARBOHYD 694 694 POTENTIAL.  
 FT CARBOHYD 754 754 POTENTIAL.  
 FT CARBOHYD 790 790 POTENTIAL.  
 FT CARBOHYD 813 813 POTENTIAL.  
 FT CARBOHYD 823 823 POTENTIAL.  
 FT CARBOHYD 851 851 POTENTIAL.  
 FT CARBOHYD 901 901 POTENTIAL.  
 FT CARBOHYD 289 291 YGS -> EFR (IN REF. 4).  
 FT CONFLICT 306 306 S -> T (IN REF. 2).  
 FT CONFLICT 520 520 P -> A (IN REF. 2).  
 FT CONFLICT 962 962 MISSING (IN REF. 4).  
 SO SEQUENCE 1025 AA: 112994 MW: DA043E62 CRC32;

Query Match 50.4%; Score 62; DB 1; Length 1025;  
 Best Local Similarity 47.8%; Pred. No. 0.1;  
 Matches 11; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Oy 1 FELVEPSISTSDNOYIGWS 23  
 Db 307 FTLIGKTNCTGSKGTGWSG 329

RESULT 6  
 DAF1\_MOUSE STANDARD; PRT; 390 AA.  
 AC Q61475; Q61397;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE COMPLEMENT DECAI-ACCELERATING FACTOR, GPI-ANCHORED PRECURSOR  
 DE (DAF-GPI).  
 GN DAF1.  
 OS MUS MUSCULUS (MOUSE).  
 OC EURAROTIA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 ON EUHERIA; RODENTIA.  
 RC [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;  
 RX MEDLINE; 95403982.  
 RA SPICER A.P., SELDIN M.F., GENDLER S.J.;  
 RL J. IMMUNOL. 155:3079-3091(1995).  
 RN [2]  
 RP SEQUENCE OF 7-390 FROM N.A.  
 RC STRAIN=BALB/C; TISSUE=SPLEEN;  
 RX MEDLINE; 96362213.  
 RA FUKUOKA Y., YASUI A., OKADA N., OKADA H.;  
 RL INT. IMMUNOL. 8:379-385(1996).  
 CC -1- FUNCTION: PROTECTION OF CELLS FROM COMPLEMENT-MEDIATED DAMAGE (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.  
 CC -1- TISSUE SPECIFICITY: BRAIN, SECRETORY EPITHELIAL, SKELETAL MUSCLE, LIVER, TESTES, THYMUS, SPLEEN AND LYMPH NODE.  
 CC -1- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE ACTIVE SITE ON SCR3 (BY SIMILARITY).  
 CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.

CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION (RCA) FAMILY.  
 CC EMBL: L41366; G886335; -  
 DR EMBL: D63679; D1010476; -  
 DR MGD: MGI:104850; DAF1.  
 KW COMPLEMENT PATHWAY; GLYCOPROTEIN; MEMBRANE; REPEAT; GPI-ANCHOR;  
 KW SIGNAL; SUSHI.  
 FT SIGNAL 1 34 POTENTIAL.  
 FT CHAIN 35 361 COMPLEMENT DECAI-ACCELERATING FACTOR, GPI-ANCHORED.  
 FT PROPER 362 390 REMOVED IN MATURE FORM (BY SIMILARITY).  
 FT DOMAIN 35 285 4 X SUSHI (SCR) REPEATS.  
 FT REPEAT 97 159 SUSHI 1.  
 FT REPEAT 162 221 SUSHI 2.  
 FT REPEAT 224 285 SUSHI 3.  
 FT DOMAIN 288 364 SUSHI 4.  
 FT DISULFID 65 94 SER/THR-RICH.  
 FT DISULFID 98 145 BY SIMILARITY.  
 FT DISULFID 129 158 BY SIMILARITY.  
 FT DISULFID 163 204 BY SIMILARITY.  
 FT DISULFID 190 220 BY SIMILARITY.  
 FT DISULFID 225 267 BY SIMILARITY.  
 FT DISULFID 253 284 BY SIMILARITY.  
 FT CARBOHYD 187 187 POTENTIAL.  
 FT CARBOHYD 262 262 POTENTIAL.  
 FT LIPID 361 361 GPI-ANCHOR (BY SIMILARITY).  
 FT CONFLICT 7 7 P -> A (IN REF. 2).  
 FT CONFLICT 9 9 T -> A (IN REF. 2).  
 FT CONFLICT 83 83 E -> G (IN REF. 2).  
 FT CONFLICT 91 91 E -> G (IN REF. 2).  
 FT CONFLICT 135 135 E -> K (IN REF. 2).  
 FT CONFLICT 173 173 H -> L (IN REF. 2).  
 FT CONFLICT 180 180 I -> T (IN REF. 2).  
 SO SEQUENCE 390 AA: 42618 MW: ADBBD34 CRC32;

Query Match 44.7%; Score 55; DB 1; Length 390;  
 Best Local Similarity 54.5%; Pred. No. 0.4;  
 Matches 12; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

Oy 1 FELVEPSISTSDNOYIGWS 22  
 Db 257 FILVGNASIVCTVSKSDVQWS 278

RESULT 7  
 INVA\_MAIZE STANDARD; PRT; 590 AA.  
 AC P49174;  
 DT 01-FEB-1996 (REL. 33, CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE BETA-FRUCTOFURANOSIDASE, CELL WALL ISOZYME PRECURSOR (EC 3.2.1.26)  
 DE (SUCROSE-6-PHOSPHATE HYDROLASE) (INVERSE).  
 OS ZEA MAIZE (MAIZE).  
 OC EURAROTIA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; MONOCOTYLEDONEAE;  
 ON CYPERALES; GRAMINEAE.  
 RC [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. BLACK MEXICAN SWEET;  
 RA SHANKER S.;  
 RL SUBMITTED (NOV-1994) TO EMBL/GENBANK/DDJ DATA BANKS.  
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL NON-REDUCING BETA-D-FRUCTOFURANOSIDE RESIDUES IN BETA-D-FRUCTOFURANOSIDES.  
 CC -1- SUBCELLULAR LOCATION: CELL WALL.  
 CC -1- INDUCTION: IN RESPONSE TO WOUNDING AND BACTERIAL INFECTION.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.  
 DR EMBL: U16995; G736359; -  
 DR MAZEDB: 113032; -  
 DR PROSITE: PS00609; GLYCOSYL\_HYDROL\_F32; FALSE NEG.  
 KW HYDROLASE; GLYCOSIDASE; GLYCOPROTEIN; CELL WALL; SIGNAL.  
 FT SIGNAL 1 28 POTENTIAL.

FT	CHAIN	29	590	BETA-FRUCTOFURANOSIDASE, CELL WALL
FT				ISOZYME.
FT	ACT_SITE	68	68	BY SIMILARITY.
FT	CARBOHYD	190	190	POTENTIAL.
FT	CARBOHYD	341	341	POTENTIAL.
SO	SEQUENCE	590 AA;	65198 MW;	B49F61E9 CRC32;

Query Match	42.3%;	Score 52;	DB 1;	Length 590;
Best Local Similarity	50.0%;	Pred. No. 1.8;		
Matches 9;	Conservative 3;	Mismatches 6;	Indels	

```
QY      6 EPSIYSTNDQVGISG 23
        11:11: 1:111
Db     112 EPAIYPSIPSDKYGCWSG 129
```

RESULT	8			
IDUGA_MOUSE				
ID	IDUGA_MOUSE	STANDARD:	PRT:	634 AA.
AC	P48441.			
DT	01-FEB-1996	(REL. 33, CREATED)		
DT	01-FEB-1996	(REL. 33, LAST SEQUENCE UPDATE)		
DT	01-FEB-1996	(REL. 33, LAST ANNOTATION UPDATE)		
DE	ALPHA-L-IDURONIDASE PRECURSOR (EC 3.2.1.76).			
GN	IDUGA.			
OS	MUS MUSCULUS (MOUSE).			
OC	EUMAROTIA, MENAZOA;	CHORDATA; VERTEBRATA; TETRAPODA, MAMMALIA		
OC	EUTHERIA; RODENTIA.			

Query Match	41.5%	Score 51	DB 1	Length 634
Best Local Similarity	56.5%	Pred. NO. 2.8		
Matches 13	Conservative 1	Mismatches 7	Indels 2	Gaps 1

```
QY      1 FELVGEPSIXSTSNDD--QVGIM 21  
          |||:||||| | | |  
Db     120 FELMGSPSGYFTDFDDKQGVFEW 142
```

RESULT	9
INV3_DAUCA	
ID	INV3_DAUCA
039693:	
AC	039693:
DT	01-NOV-1997 (REL. 35, CREATED)
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE	BETA-FRUCTOFURANOSIDASE, INSOLUBLE ISOENZYME 3 PRECURSOR (EC 3.2.1.26)
DE	(SUCROSE-6-PHOSPHATE HYDROLASE 3) (INVERTASE 3) (CELL WALL BETA-FRUCTOSIDASE 3).
DE	INV3.
GN	DAUCUS CAROTA (CARROT).
OS	DAUCUS CAROTA (CARROT).

CC	EUCHARYOTA; PLANTIA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE; APiales		
OC	UMELLIFERAE.		
CC	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-CV, QUEEN ANNE'S LACE;		
RX	MEDLINE; 95306787.		
CC	LORENZ K., LIENHARD S., STUEN A.;		
RL	PLANT MOL. BIOL. 28:189-194(1995).		
CC	-1- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN PHLOEM UNLOADING AND IN STRESS RESPONSE.		
CC	-1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL NON-REDUCING BETA-D-FUCOPYRANOSIDE RESIDUES IN BETA-D-PROCTOFURANOSIDES.		
CC	-1- SUBCELLULAR LOCATION: IONICALLY BOUND TO THE CELL WALL.		
CC	-1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.		
DR	EMBL, X78423; GI020102;		
DR	PROSITE; PS00609; GLYCOSTYL-HYDROL_F32. 1.		
KW	HYDROLASE; GLYCOSIDASE; GLYCOPROTEIN; CELL WALL; ZYMOGEN; SIGNAL.		
FT	FT SIGNAL	1	POTENTIAL.
FT	FT PROPEP	?	POTENTIAL.
FT	CHAIN	?	BETA-FRUCTOFURANOSIDASE, INSOLUBLE
FT		583	ISOENZYME 3.
FT	CARBOHYD	164	POTENTIAL.
FT	CARBOHYD	280	POTENTIAL.
FT	CARBOHYD	303	POTENTIAL.
FT	CARBOHYD	340	POTENTIAL.
FT	CARBOHYD	561	POTENTIAL.
SO	SEQUENCE	583 AA;	66381 MW; DF188EED CRC32;

Query Match	41.5%	Score 51;	DB 1;	Length 583;
Best Local Similarity	50.0%	Prod. No. 2.6;		
Matches	9;	Conservative	3;	Mismatches 6;
				Indels 0;
				Gaps 0

```

OY      6 EPSIYSTNDQVGISG 23
         ||:::  ||| |||
Db      112 EPAIFPSKPFQYGCWSG 129

```

RESULT	10
CABP_HUMAN	
ID	CABP_HUMAN
AC	P04003;
DT	23-OCT-1986 (REL. 02, CREATED)
DT	01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE	CAB-BINDING PROTEIN ALPHA CHAIN PRECURSOR (PROLINE-RICH PROTEIN) (PRP).
DE	CABPA OR CABP.
GN	HOMO SAPIENS (HUMAN).
OS	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC	EUTHERIA; PRIMATES.
CC	[1]
RN	SEQUENCE FROM N.A.
RP	TISSUE=LIVER;
RC	MEDLINE: 90073699.
RX	MATSUGUCHI T., OKAMURA S., ASO T., SATA T., NITHO Y.;
RL	BIOCHEM. BIOPHYS. RES. COMMUN. 165:138-144(1989).
RL	[2]
RP	SEQUENCE FROM N.A.
RC	MEDLINE: 91113199.
RX	ASO T., OKAMURA S., MATSUGUCHI T., SAKAMOTO N., SATA T., NITHO Y.;
RL	BIOCHEM. BIOPHYS. RES. COMMUN. 174:222-227(1991).
RL	[3]
RN	SEQUENCE OF 9-81 FROM N.A.
RP	MEDLINE: 88242821.
RX	LINTIN S.J., LEWIN A.R., REID K.B.M.;
RL	FEBS LETT. 232:328-332(1988).
RL	[4]
RN	SEQUENCE OF 203-288 FROM N.A.
RP	MEDLINE: 86301119.
RX	LINTIN S.J., REID K.B.M.;
RL	FEBS LETT. 204:77-81(1986).
RL	[5]

RP SEQUENCE OF 80-597 FROM N.A.  
 RX MEDLINE; 86025405.  
 RA CHUNG L.P., BENTLEY D.R., REID K.B.M.;  
 RL BIOCHEM. J. 230:133-141(1985).  
 RN [6]  
 RP SEQUENCE OF 49-88.  
 RX MEDLINE; 85296001.  
 RA CHUNG L.P., GAGNON J., REID K.B.M.;  
 RL MOL. IMMUNOL. 22:427-435(1985).  
 RN [7]  
 RP ELECTRON MICROSCOPY, 3-DIMENSIONAL STRUCTURE, AND LIGAND BINDING.  
 RA DAHLBACK B., SMITH C.A., MUELLER-EBERHARD H.J.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 80:3461-3465(1983).  
 CC -1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT  
 ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR  
 (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT  
 ALSO ACCELERATES THE DEGRADATION OF THE C4B2A COMPLEX (C3  
 CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA  
 CHAIN BINDS C4B. IT INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S  
 AND WITH SERUM AMYLOID P COMPONENT.  
 CC -1- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS  
 OF 3 POSSIBLE SORTS: A 570 KD COMPLEX OF 7 ALPHA CHAINS AND 1 BETA  
 CHAIN, A 530 KD HOMOPENTAMER OF ALPHA CHAINS OR A 500 KD COMPLEX  
 OF 6 ALPHA CHAINS AND 1 BETA CHAIN. THE CENTRAL BODY OF THE  
 ALPHA CHAIN HOMOPOLYMER SUPPORTS TENTACLES, EACH WITH THE BINDING  
 SITE FOR C4B AT THE END.  
 CC -1- TISSUE SPECIFICITY: CHYLOMICRONS IN THE PLASMA.  
 CC -1- SIMILARITY: CONTAINS 8 SUSHI (SCR) REPEATS.  
 CC -1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.  
 DR EMBL; M31452; G190502; -  
 DR EMBL; M62486; G190500; JOINED.  
 DR EMBL; M62475; G190500; JOINED.  
 DR EMBL; M62476; G190500; JOINED.  
 DR EMBL; M62477; G190500; JOINED.  
 DR EMBL; M62478; G190500; JOINED.  
 DR EMBL; M62479; G190500; JOINED.  
 DR EMBL; M62480; G190500; JOINED.  
 DR EMBL; M62481; G190500; JOINED.  
 DR EMBL; M62482; G190500; JOINED.  
 DR EMBL; M62484; G190500; JOINED.  
 DR EMBL; M62485; G190500; JOINED.  
 DR EMBL; X07853; E3792; -  
 DR EMBL; X04284; E12637; -  
 DR EMBL; X04296; E12638; -  
 DR EMBL; X02865; G29565; -  
 DR PIR; A33568; NBHUC4.  
 DR HSP; P08603; IHCC.  
 DR MIM; 120830; -  
 KW COMPLEMENT PATHWAY; PLASMA; GLYCOPROTEIN; REPEAT; SUSHI; SIGNAL;  
 KM POLYMORPHISM.  
 FT SIGNAL 1 48  
 FT CHAIN 49 597 C4B-BINDING PROTEIN ALPHA CHAIN.  
 FT DOMAIN 49 539 8 X SUSHI (SCR) REPEATS.  
 FT REPEAT 49 109 SUSHI 1.  
 FT REPEAT 112 171 SUSHI 2.  
 FT REPEAT 174 235 SUSHI 3.  
 FT REPEAT 238 295 SUSHI 4.  
 FT REPEAT 298 361 SUSHI 5.  
 FT REPEAT 364 423 SUSHI 6.  
 FT REPEAT 425 481 SUSHI 7.  
 FT REPEAT 483 539 SUSHI 8.  
 FT DISULFID 50 96 BY SIMILARITY.  
 FT DISULFID 81 108 BY SIMILARITY.  
 FT DISULFID 113 154 BY SIMILARITY.  
 FT DISULFID 140 170 BY SIMILARITY.  
 FT DISULFID 175 217 BY SIMILARITY.  
 FT DISULFID 203 234 BY SIMILARITY.  
 FT DISULFID 239 281 BY SIMILARITY.  
 FT DISULFID 267 294 BY SIMILARITY.  
 FT DISULFID 299 348 BY SIMILARITY.  
 FT DISULFID 332 360 BY SIMILARITY.

FT DISULFID 7364 7387 BY SIMILARITY.  
 FT DISULFID 7365 7409 BY SIMILARITY.  
 FT DISULFID 7399 7422 BY SIMILARITY.  
 FT DISULFID 426 468 BY SIMILARITY.  
 FT DISULFID 454 480 BY SIMILARITY.  
 FT DISULFID 484 525 BY SIMILARITY.  
 FT DISULFID 511 538 BY SIMILARITY.  
 FT DISULFID 546 546 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).  
 FT DISULFID 558 558 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).  
 FT CARBOHYD 506 506 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).  
 FT CARBOHYD 528 528  
 FT CARBOHYD 92 92  
 FT VARIANT 357 357  
 FT VARIANT 357 357  
 SQ SEQUENCE 597 AA; 67033 MW; DA7BADCS CRC32;  
 Q -> T.  
 Y -> H.  
 Query Match 39.88; Score 49; DB 1; Length 597;  
 Best Local Similarity 38.14; Pred. No. 5.4;  
 Matches 8; Conservative 5; Mismatches 8; Indels 0; Gaps 0;  
 Oy 1 FELYGEPSTYSNDPOGVW 21  
 Db 207 FSLGHASICTVETITGVW 227  
 RESULT 11  
 C4BP-RAT STANDARD; PRT; 558 AA.  
 ID C4BP-RAT  
 AC Q63514;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
 DE C4B-BINDING PROTEIN ALPHA CHAIN PRECURSOR.  
 GN CABPA.  
 OS RATUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-LIVER.  
 RX MEDLINE; 97166082.  
 RL HILLARP A., WIKLUND H., TERNER A., DAHLBACK B.;  
 RL J. IMMUNOL. 158:1315-1323(1997).  
 CC -1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT  
 ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR  
 (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT  
 ALSO ACCELERATES THE DEGRADATION OF THE C4B2A COMPLEX (C3  
 CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA  
 CHAIN BINDS C4B. IT INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S  
 AND WITH SERUM AMYLOID P COMPONENT.  
 CC -1- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS.  
 CC -1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.  
 CC -1- SIMILARITY: CONTAINS 8 SUSHI (SCR) REPEATS.  
 DR EMBL; Z50051; G699380; -  
 KW COMPLEMENT PATHWAY; PLASMA; GLYCOPROTEIN; REPEAT; SUSHI; SIGNAL;  
 KM COMPLEMENT PATHWAY; PLASMA; GLYCOPROTEIN; REPEAT; SUSHI; SIGNAL;  
 FT SIGNAL 1 13  
 FT CHAIN 14 558 C4B-BINDING PROTEIN ALPHA CHAIN.  
 FT DOMAIN 14 502 8 X SUSHI (SCR) REPEATS.  
 FT REPEAT 14 73 SUSHI 1.  
 FT REPEAT 14 73 SUSHI 1.  
 FT REPEAT 138 200 SUSHI 2.  
 FT REPEAT 203 259 SUSHI 3.  
 FT REPEAT 262 325 SUSHI 4.  
 FT REPEAT 328 387 SUSHI 5.  
 FT REPEAT 389 444 SUSHI 6.  
 FT REPEAT 446 502 SUSHI 7.  
 FT REPEAT 446 502 SUSHI 8.  
 FT DISULFID 15 60 BY SIMILARITY.  
 FT DISULFID 45 72 BY SIMILARITY.  
 FT DISULFID 77 118 BY SIMILARITY.  
 FT DISULFID 104 134 BY SIMILARITY.  
 FT DISULFID 139 182 BY SIMILARITY.  
 FT DISULFID 168 199 BY SIMILARITY.

FI REPEAT 16/ 226 SUSHI 3.

Db 119 EPAIFPSKPFDKYGCWSG 136

Db 119 EPAIFPSKPFEDKYGCWSG 136

```

RESULT 14
RN CAD3_HUMAN STANDARD; PRT; 829 AA.
ID CAD3_HUMAN
AC P22223;
DT 01-AUG-1991 (REL. 19, CREATED)
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE PLACENTAL-CADHERIN PRECURSOR (P-CADHERIN).
GN CDH3 OR CDHF.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90009051.
RA SHIMOYAMA Y., YOSHIDA T., TERADA M., SHIMOSATO Y., ABE O.,
RA HIROASHI S.; 109:1787-1794(1989).
RL J. CELL. BIOL. 109:1787-1794(1989).
CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
DR EMBL: X63629; G35323; -.
DR PIR: A33659; IJHUCP.
DR MIM: 114021.
DR PROSITE: PS00232; CADHERIN: 3.
KW CELL ADHESION; GLYCOPROTEIN; PHOSPHORYLATION; TRANSMEMBRANE;
KW CALCULUM-BINDING; REPEAT; SIGNAL.
FT SIGNAL 1 26
FT PROPEP 27 107 P-CADHERIN.
FT CHAIN 108 829 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 108 654 POTENTIAL.
FT TRANSMEM 655 677 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 678 829 POTENTIAL.
FT REPEAT 108 215 CADHERIN 1.
FT REPEAT 216 328 CADHERIN 2.
FT REPEAT 329 440 CADHERIN 3.
FT REPEAT 441 546 CADHERIN 4.
FT REPEAT 547 650 CADHERIN 5.
FT DOMAIN 785 800 SER-RICH.
FT CARBOHYD 200 200 POTENTIAL.
FT CARBOHYD 566 566 POTENTIAL.
SO SEQUENCE 829 AA; 91427 MW; A42E08FC CRC32;

Query Match 36.6%; Score 45; DB 1; Length 829;
Best Local Similarity 63.6%; Pred. No. 32;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 6 EPISTYNSND 16
DB 62 EPALFSTDND 72

```

```

RESULT 15
RN D3DR_MOUSE STANDARD; PRT; 446 AA.
ID D3DR_MOUSE
AC P30728;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE D(3) DOPAMINE RECEPTOR.
GN DRD3.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-BALB/C;
RA FISHBURN C.S., DAVID C., CARMON S., BELLELI D., FUCHS S.;

RL SUBMITTED (JUL-1992) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95322026.
RA FU D., SKRABIN B. V., BROSTIUS J., ROBAKIS N. K.;
RL DNA CELL BIOL. 14:485-492(1995).
CC -1- FUNCTION: THIS IS ONE OF THE FIVE TYPES (D1 TO D5) OF RECEPTORS
CC FOR DOPAMINE. THE ACTIVITY OF THIS RECEPTOR IS NOT MEDIATED BY G
CC PROTEINS WHICH INTERACT WITH ADENYLYL CYCLASE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF THIS PROTEIN ARE PRODUCED BY
CC ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR EMBL: X67274; G50652; -.
DR EMBL: S78580; G1042108; JOINED.
DR EMBL: S78574; G1042108; JOINED.
DR EMBL: S78638; G1042108; JOINED.
DR EMBL: S78577; G1042108; JOINED.
DR EMBL: S78578; G1042108; JOINED.
DR EMBL: S78579; G1042108; JOINED.
DR EMBL: S78639; G1042108; JOINED.
DR PIR: S25161; S25161.
DR GCRDB: GCR_0449; -.
DR MGD; MG1:94925; DRD3.
DR PROSITE: PS00237; G_PROTEIN_RECEPTOR; 1.
DR G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
KW MULTIGENE FAMILY; ALTERNATIVE SPLICING.
FT DOMAIN 1 32 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 33 55 1 (POTENTIAL).
FT DOMAIN 56 66 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 67 92 2 (POTENTIAL).
FT DOMAIN 93 104 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 105 126 3 (POTENTIAL).
FT DOMAIN 127 149 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 150 172 4 (POTENTIAL).
FT DOMAIN 173 185 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 186 209 5 (POTENTIAL).
FT DOMAIN 210 375 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 376 397 6 (POTENTIAL).
FT DOMAIN 398 412 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 413 434 7 (POTENTIAL).
FT DOMAIN 435 446 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 12 12 POTENTIAL.
FT CARBOHYD 19 19 POTENTIAL.
FT CARBOHYD 97 97 POTENTIAL.
FT CARBOHYD 173 173 POTENTIAL.
FT DISULFID 103 181 BY SIMILARITY.
FT VARSPLIC 268 288 MISSING (IN D3SHORT).
SO SEQUENCE 446 AA; 49669 MW; 77A65F3C CRC32;

Query Match 36.6%; Score 45; DB 1; Length 446;
Best Local Similarity 55.6%; Pred. No. 16;
Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
QY 1 FELVGEPSIYSTNSNDQV 18
DB 172 FNTTGPDSICISINPDFV 189

```

Search completed: July 18, 1999, 11:37:54  
Job time: 1753 sec

CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.  
 CC -1- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR  
 CC FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE  
 CC ACTIVE SITE ON SCR3 (BY SIMILARITY).  
 CC -1- PTM: THE SER/THR-RICH DOMAIN IS HEAVILY O-GLYCOSYLATED.  
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF DAF (DAF-2, SHOWN HERE, AND  
 CC DAF-1) ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.  
 CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.  
 CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION  
 CC (RCA) FAMILY.  
 DR EMBL: S67775; G459390; -  
 KW COMPLEMENT PATHWAY; PLASMA; GLYCOPROTEIN; MEMBRANE; REPEAT;  
 KM ALTERNATIVE SPLICING; GPI-ANCHOR; SUSHI.  
 FT NON\_TER 1 1  
 FT CHAIN <1 312 COMPLEMENT DEGRADATION-ACCELERATING FACTOR.  
 FT PROPEP 313 340 REMOVED IN NATURE FORM (BY SIMILARITY).  
 FT DOMAIN <1 243 4 X SUSHI (SCR) REPEATS.  
 FT REPEAT <1 54 SUSHI 1.  
 FT REPEAT 56 118 SUSHI 2.  
 FT REPEAT 121 180 SUSHI 3.  
 FT REPEAT 183 243 SUSHI 4.  
 FT DOMAIN 246 315 SER/THR-RICH.  
 FT DISULFID 24 53 BY SIMILARITY.  
 FT DISULFID 57 104 BY SIMILARITY.  
 FT DISULFID 88 117 BY SIMILARITY.  
 FT DISULFID 122 163 BY SIMILARITY.  
 FT DISULFID 149 179 BY SIMILARITY.  
 FT DISULFID 184 226 BY SIMILARITY.  
 FT DISULFID 212 242 BY SIMILARITY.  
 FT CARBOHYD 54 54 POTENTIAL.  
 FT CARBOHYD 107 107 POTENTIAL.  
 FT LIPID 312 312 GPI-ANCHOR (BY SIMILARITY).  
 SQ SEQUENCE 340 AA; 37180 MW; 5CACC18F8 CRC32;

Query Match 54.18; Score 66.5; DB 1; Length 340;  
 Best Local Similarity 56.58; Pred. No. 0.0059;  
 Matches 13; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

QY 1 FELVGEISYSTNDQVIGMSG 23  
 DB 216 YTMGEHSYCTVNDDE-GEWSG 237

RESULT 4  
 DAF\_HUMAN STANDARD; PTI; 361 AA.  
 AC P08174; P08679;  
 DT 01-AUG-1988 (REL. 08, CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE COMPLEMENT DEGRADATION-ACCELERATING FACTOR PRECURSOR (CD55).  
 OS DAF OR CD55.  
 GN HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 87115845.  
 RA CARAS I.W., DAYTZ M.A., RHEE L., WEDDELL G., MARTIN D.W. JR.,  
 RA NUSSENZWEIG V.;  
 RL NATURE 325:545-549(1987).  
 RN [2]  
 RP SEQUENCE OF 6-381 FROM N.A. (DAF-2 FORM).  
 RX MEDLINE: 87175602.  
 RA MEDOF M.E., LUBLIN D.M., HOLERS V.M., AYERS D.J., GETTY R.R.,  
 RA LEVKAM J.F., ATKINSON J.P., TYKOCINSKI M.L.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 84:2007-2011(1987).  
 RN [3]  
 RP SEQUENCE OF 1-100 FROM N.A.  
 RX MEDLINE: 91271256.  
 RA EWOLONU U.K., RAVI L., MEDOF M.E.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 88:4675-4679(1991).

RN [4]  
 RP SEQUENCE OF 35-46.  
 RC TISSUE-ORIGIN:  
 RX MEDLINE: 91291869.  
 RA NAKANO Y., SUGITA Y., ISHIKAWA Y., CHOI N.-H., TOBE T., TOMITA M.;  
 RL BIOCHIM. BIOPHYS. ACTA 1074:326-330(1991).  
 RN [5]  
 RP GPI-ANCHOR.  
 RX MEDLINE: 91093238.  
 RA MORAN P., RAAB H., KOHR W.J., CARAS I.W.;  
 RL J. BIOL. CHEM. 266:1250-1257(1991).  
 RN [6]  
 RP DISULFIDE BONDS IN SUSHI DOMAINS.  
 RX MEDLINE: 92305034.  
 RA NAKANO Y., SUMIDA K., KIKUTA N., MURA N.-H., TOBE T., TOMITA M.;  
 RL BIOCHIM. BIOPHYS. ACTA 1116:235-240(1992).  
 RN [7]  
 RP FUNCTION AS A ECHOVIRUS RECEPTOR.  
 RX MEDLINE: 95045399.  
 RA WARD T., PIRKIN P.A., CLARKSON N.A., STONE D.M., MINOR P.D.,  
 RA ALMOND J.W.;  
 RL EMBO J. 13:5070-5074(1994).  
 RN [8]  
 RP VARIANT BLOOD GROUP DR(A-).  
 RX MEDLINE: 94325573.  
 RA LUBLIN D.M., MALLINSON G., POOLE J., REID M.E., THOMPSON E.S.,  
 RA PERDAN B.R., TELER M.J., ANSTER D.J., TANNER M.J.A.;  
 RL BLOOD 84:1276-1282(1994).  
 CC -1- FUNCTION: THIS PROTEIN RECOGNIZES C4B AND C3B FRAGMENTS THAT  
 CC CONDENSE WITH CELL-SURFACE HYDROXYL OR AMINO GROUPS WHEN NASCENT  
 CC C4B AND C3B ARE LOCALLY GENERATED DURING C4 AND C3 ACTIVATION.  
 CC INTERACTION OF DAF WITH CELL-ASSOCIATED C4B AND C3b POLYPEPTIDES  
 CC INTERFERES WITH THEIR ABILITY TO CATALYZE THE CONVERSION OF C2 AND  
 CC FACTOR B TO ENZYMATICALLY ACTIVE C2a AND Bb AND THEREBY PREVENTS  
 CC THE FORMATION OF C4b2a AND C3bBb, THE AMPLIFICATION CONVERTASES OF  
 CC THE COMPLEMENT CASCADE.  
 CC -1- FUNCTION: ALSO ACTS AS THE RECEPTOR FOR ECHOVIRUS 7 AND RELATED  
 CC VIRUSES (ECHOVIRUSES 13, 21, 29 AND 33).  
 CC -1- SUBUNIT: MONOMER (MAJOR FORM) AND NON-DISULFIDE-LINKED, COVALENT  
 CC HOMODIMER (MINOR FORM).  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED ON THE PLASMA MEMBRANES OF ALL CELL  
 CC TYPES THAT ARE IN INTIMATE CONTACT WITH PLASMA COMPLEMENT  
 CC PROTEINS. IT IS ALSO FOUND ON THE SURFACES OF EPITHELIAL CELLS  
 CC LINING EXTRACELLULAR COMPARTMENTS, AND VARIANTS OF THE MOLECULE  
 CC ARE PRESENT IN BODY FLUIDS AND IN EXTRACELLULAR MATRIX.  
 CC -1- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR  
 CC FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE  
 CC ACTIVE SITE ON SCR3 (BY SIMILARITY).  
 CC -1- POLYMORPHISM: DAF IS RESPONSIBLE FOR THE CROHER BLOOD GROUP  
 CC SYSTEM. IT CONSISTS OF AT LEAST SEVEN HIGH-INCIDENCE (CRA),  
 CC TC(A), DR(A), ES(A), MES(B), DMC, AND IFC) AND LOW-INCIDENCE  
 CC TC(B), TC(C), AND MES(A)) ANTIGENS THAT RESIDE ON DAF. IN THE  
 CC CROHER PHENOTYPES DR(A-) AND INAB THERE IS REDUCED OR ABSENT  
 CC EXPRESSION OF DAF, RESPECTIVELY. IN THE CASE OF THE DR(A-)  
 CC PHENOTYPE, A SINGLE NUCLEOTIDE SUBSTITUTION WITHIN EXON 5 ACCOUNTS  
 CC FOR TWO CHANGES: A SIMPLE AMINO ACID SUBSTITUTION THAT IS THE  
 CC BASIS OF THE ANTIGENIC VARIATION, AND AN ALTERNATIVE SPLICING  
 CC EVENT THAT UNDERLIES THE DECREASED EXPRESSION OF DAF IN THIS  
 CC PHENOTYPE.  
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF DAF (DAF-2, SHOWN HERE, AND  
 CC DAF-1) ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.  
 CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.  
 CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION  
 CC (RCA) FAMILY.  
 CC -1- DATABASE: NAME-PROV: NOTE-CD guide CD55 entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/row/cd/cd55.htm".  
 DR EMBL: M31516; G181468; -  
 DR EMBL: M30142; G181465; -  
 DR EMBL: M15799; G181463; -  
 DR EMBL: M64653; G181476; -  
 DR EMBL: M64356; G181476; JOINED.

Jul 19 13:55:31 1999

RL J. BIOL. CHEM. 264:2118-2125(1989).  
 RN [2]  
 RP SEQUENCE OF 226-233; 256-267; 332-341; 667-677 AND 898-908.  
 RX MEDLINE: 86287311.  
 RA WEIS J.J., FEARON D.T., KLICKSTEIN L.B., MONG W.W., RICHARDS S.A.,  
 RA DE BRUYN KOS A., SMITH J.A., WEIS J.H.;  
 RL PROC. NATL. ACADE. SCI. U.S.A. 83:5639-5643(1986).  
 CC -1- FUNCTION: RECEPTOR FOR COMPLEMENT C3D AND FOR THE EPSTEIN-BARR  
 CC VIRUS ON HUMAN B-CELLS. PARTICIPATES IN B LYMPHOCYTES ACTIVATION.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: MATURE B LYMPHOCYTES AND FOLLICULAR  
 CC DENDRITIC CELLS OF THE SPLEEN.  
 CC -1- SIMILARITY: TO MOUSE CR2. CD21 IS A MEMBER OF THE REGULATORS OF  
 CC COMPLEMENT ACTIVATION (RCA) FAMILY.  
 CC -1- SIMILARITY: CONTAINS 15 SUSHI (SCR) REPEATS.  
 CC -1- DATABASE: NAME-PRO: NOTE-CD guide cd21 entry;  
 CC WWW=<http://www.ncbi.nlm.nih.gov/prov/cd/cd21.htm>.  
 CC  
 DR EMBL: M26004; G181940; -.  
 DR EMBL: M26016; G181938; -.  
 DR EMBL: M24007; G181938; JOINED.  
 DR EMBL: M24008; G181938; JOINED.  
 DR EMBL: M24009; G181938; JOINED.  
 DR EMBL: M24010; G181938; JOINED.  
 DR EMBL: M24011; G181938; JOINED.  
 DR EMBL: M26009; G181938; JOINED.  
 DR EMBL: M26011; G181938; JOINED.  
 DR EMBL: M26012; G181938; JOINED.  
 DR EMBL: M26013; G181938; JOINED.  
 DR EMBL: M26014; G181938; JOINED.  
 DR EMBL: M26015; G181938; JOINED.  
 DR PIR: A32036; A32036.  
 DR PIR: A24319; A24319.  
 DR PIR: B24319; B24319.  
 DR PIR: C24319; C24319.  
 DR PIR: D24319; D24319.  
 DR PIR: E24319; E24319.  
 DR PIR: F24319; F24319.  
 DR HSSP: P08603; 1HFI.  
 DR MIM: 120650; -.  
 KW COMPLEMENT PATHWAY; GLYCOPROTEIN; TRANSMEMBRANE; REPEAT; SIGNAL;  
 KW RECEPTOR; SUSHI.  
 FT  
 FT SIGNAL 1 20  
 FT CHAIN 21 1033  
 FT DOMAIN 21 971  
 FT TRANSMEM 972 999  
 FT DOMAIN 1000 1033  
 FT DOMAIN 22 971  
 FT REPEAT 22 83  
 FT REPEAT 90 147  
 FT REPEAT 153 211  
 FT REPEAT 214 272  
 FT REPEAT 275 343  
 FT REPEAT 350 407  
 FT REPEAT 409 467  
 FT REPEAT 470 523  
 FT REPEAT 526 594  
 FT REPEAT 601 658  
 FT REPEAT 661 715  
 FT REPEAT 718 780  
 FT REPEAT 787 844  
 FT REPEAT 850 908  
 FT REPEAT 911 969  
 FT DISULFID 23 65  
 FT DISULFID 51 82  
 FT DISULFID 91 132  
 FT DISULFID 118 146  
 FT DISULFID 154 197  
 FT DISULFID 183 210  
 FT DISULFID 215 256  
 FT DISULFID 242 271  
 FT DISULFID 276 325  
 FT DISULFID 305 342

FT DISULFID 351 393 BY SIMILARITY.  
 FT DISULFID 379 406 BY SIMILARITY.  
 FT DISULFID 410 453 BY SIMILARITY.  
 FT DISULFID 439 466 BY SIMILARITY.  
 FT DISULFID 471 509 BY SIMILARITY.  
 FT DISULFID 495 522 BY SIMILARITY.  
 FT DISULFID 527 576 BY SIMILARITY.  
 FT DISULFID 556 593 BY SIMILARITY.  
 FT DISULFID 602 644 BY SIMILARITY.  
 FT DISULFID 630 657 BY SIMILARITY.  
 FT DISULFID 662 699 BY SIMILARITY.  
 FT DISULFID 685 714 BY SIMILARITY.  
 FT DISULFID 719 762 BY SIMILARITY.  
 FT DISULFID 748 779 BY SIMILARITY.  
 FT DISULFID 788 830 BY SIMILARITY.  
 FT DISULFID 816 843 BY SIMILARITY.  
 FT DISULFID 851 894 BY SIMILARITY.  
 FT DISULFID 880 907 BY SIMILARITY.  
 FT DISULFID 912 955 BY SIMILARITY.  
 FT DISULFID 941 968 BY SIMILARITY.  
 FT CARBOHYD 121 121 POTENTIAL.  
 FT CARBOHYD 127 127 POTENTIAL.  
 FT CARBOHYD 294 294 POTENTIAL.  
 FT CARBOHYD 372 372 POTENTIAL.  
 FT CARBOHYD 492 492 POTENTIAL.  
 FT CARBOHYD 623 623 POTENTIAL.  
 FT CARBOHYD 682 682 POTENTIAL.  
 FT CARBOHYD 800 800 POTENTIAL.  
 FT CARBOHYD 823 823 POTENTIAL.  
 FT CARBOHYD 861 861 POTENTIAL.  
 FT CARBOHYD 911 911 POTENTIAL.  
 FT CONFLICT 667 667 O -> D (IN REF. 2).  
 FT CONFLICT 902 902 O -> G (IN REF. 2).  
 FT CONFLICT 906 906 H -> L (IN REF. 2).  
 SQ SEQUENCE 1033 AA: 112973 MW: 70325558 CRC32:

Query Match 57.7%; Score 71; DB 1; Length 1033;  
 Best Local Similarity 56.5%; Pred. No. 0.0042;  
 Matches 13; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 FELVEPSTYSTSNDQVGIMSG 23  
 Db 566 FSLIGESTIRCTSDQERGTMWG 588  
 ID DAF\_PONPY STANDARD; PRT; 340 AA.  
 AC P49457;  
 DT 01-FEB-1996 (REL. 33, CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE COMPLEMENT DECAY-ACCELERATING FACTOR (CD55) (FRAGMENT).  
 GN DAF OR CD55.  
 OS PONGO PYGMAEUS (ORANGUTAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX NICKELS M.W., ALVAREZ J.I., LUBLIN D.M., ATKINSON J.P.;  
 RL J. IMMUNOL. 152:676-685(1994).  
 CC -1- FUNCTION: THIS PROTEIN RECOGNIZES C4B AND C3B FRAGMENTS THAT  
 CC CONDENSE WITH CELL-SURFACE HYDROXYL OR AMINO GROUPS WHEN NASCENT  
 CC C4B AND C3B ARE LOCALLY GENERATED DURING C4 AND C3 ACTIVATION.  
 CC INTERFERES WITH THEIR ABILITY TO CATALYZE THE CONVERSION OF C2 AND  
 CC FACTOR B TO ENZYMATICALLY ACTIVE C2A AND BB AND THEREBY PREVENTS  
 CC THE FORMATION OF C4B2A AND C3BB, THE AMPLIFICATION CONVERTASES OF  
 CC THE COMPLEMENT CASCADE (BY SIMILARITY).  
 CC -1- SUBUNIT: MONOMER (MAJOR FORM) AND NON-DISULFIDE-LINKED, COVALENT  
 CC HOMODIMER (MINOR FORM).





DR EMBL: L17412; G306680; JOINED.  
 DR EMBL: L17413; G306680; JOINED.  
 DR EMBL: L17414; G306680; JOINED.  
 DR EMBL: L17415; G306680; JOINED.  
 DR PFAM: PF00084; sushi; 30.  
 SQ SEQUENCE 2039 AA; 223603 MW; A4972215 CRC32;

Query Match 95.9%; Score 118; DB 2; Length 2039;  
 Best Local Similarity 95.7%; Pred. No. 5.7e-10;  
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FELVGEPSIYTSNDQVGIMSG 23  
 |||||  
 DB 205 FELVGEPSIYCTSDNDQVGIMSG 227

RESULT 2

ID 016744 PRELIMINARY; PRT; 2489 AA.

AC 016744;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE COMPLEMENT RECEPTOR 1.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
 OC CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 94065175.  
 RA VIR D.P., WONG W.W.;  
 RT "Structure of the gene for the F allele of complement receptor type 1  
 and sequence of the coding region unique to the S allele.";  
 RL J. IMMUNOL. 151:6214-6224(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA VIR D.P., WONG W.W.;  
 RL SUBMITTED (JUN-1993) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL: L17399; G451303; JOINED.  
 DR EMBL: L17409; G451303; JOINED.  
 DR EMBL: L17419; G451303; JOINED.  
 DR EMBL: L17420; G451303; JOINED.  
 DR EMBL: L17421; G451303; JOINED.  
 DR EMBL: L17422; G451303; JOINED.  
 DR EMBL: L17423; G451303; JOINED.  
 DR EMBL: L17424; G451303; JOINED.  
 DR EMBL: L17425; G451303; JOINED.  
 DR EMBL: L17426; G451303; JOINED.  
 DR EMBL: L17427; G451303; JOINED.  
 DR EMBL: L17428; G451303; JOINED.  
 DR EMBL: L17429; G451303; JOINED.  
 DR EMBL: L17430; G451303; JOINED.  
 DR EMBL: L17391; G451303; JOINED.  
 DR EMBL: L17392; G451303; JOINED.  
 DR EMBL: L17393; G451303; JOINED.  
 DR EMBL: L17394; G451303; JOINED.  
 DR EMBL: L17418; G451303; JOINED.  
 DR EMBL: L17390; G451303; JOINED.  
 DR EMBL: L17395; G451303; JOINED.  
 DR EMBL: L17416; G451303; JOINED.  
 DR EMBL: L17417; G451303; JOINED.  
 DR EMBL: L17396; G451303; JOINED.  
 DR EMBL: L17397; G451303; JOINED.  
 DR EMBL: L17398; G451303; JOINED.  
 DR EMBL: L17400; G451303; JOINED.  
 DR EMBL: L17401; G451303; JOINED.  
 DR EMBL: L17402; G451303; JOINED.  
 DR EMBL: L17403; G451303; JOINED.  
 DR EMBL: L17405; G451303; JOINED.  
 DR EMBL: L17406; G451303; JOINED.  
 DR EMBL: L17407; G451303; JOINED.  
 DR EMBL: L17408; G451303; JOINED.

DR EMBL: L17410; G451303; JOINED.  
 DR EMBL: L17411; G451303; JOINED.  
 DR EMBL: L17412; G451303; JOINED.  
 DR EMBL: L17413; G451303; JOINED.  
 DR EMBL: L17414; G451303; JOINED.  
 DR EMBL: L17415; G451303; JOINED.  
 DR PFAM: PF00084; sushi; 37.  
 SQ SEQUENCE 2489 AA; 272846 MW; 5869B6F9 CRC32;

Query Match 95.9%; Score 118; DB 2; Length 2489;  
 Best Local Similarity 95.7%; Pred. No. 7.2e-10;  
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FELVGEPSIYTSNDQVGIMSG 23  
 |||||  
 DB 205 FELVGEPSIYCTSDNDQVGIMSG 227

RESULT 3

ID 029531 PRELIMINARY; PRT; 661 AA.

AC 029531;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE COMPLEMENT RECEPTOR 1 PRECURSOR (FRAGMENT).  
 GN CRL  
 OS PAN TROGLODYTES (CHIMPANZEE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
 OC CATARRHINI; HOMINIDAE; PAN.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 94292799.  
 RA BIRNINSHAM D.J., SHEN X.P., HOURCADE D., NICKELLS M.W., ATKINSON J.P.;  
 RT "Primary sequence of an alternatively spliced form of CRL. Candidate  
 for the 75,000 M(r) complement receptor expressed on chimpanzee  
 erythrocytes.";  
 RL J. IMMUNOL. 153:691-700(1994).  
 DR EMBL: L24921; G557727; 9.  
 DR PFAM: PF00084; sushi; 9.  
 DR SIGNAL; ALTERNATIVE SPLICING.  
 FT NON\_TER 1  
 FT SIGNAL <1 16  
 FT CHAIN 17 >661  
 FT NON\_TER 661  
 SQ SEQUENCE 661 AA; 72966 MW; 9D78E262 CRC32;

Query Match 95.9%; Score 118; DB 4; Length 661;  
 Best Local Similarity 95.7%; Pred. No. 1.6e-10;  
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FELVGEPSIYTSNDQVGIMSG 23  
 |||||  
 DB 180 FELVGEPSIYCTSDNDQVGIMSG 202

RESULT 4

ID 029530 PRELIMINARY; PRT; 2014 AA.

AC 029530;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE COMPLEMENT RECEPTOR 1 (FRAGMENT).  
 GN CRL.  
 OS PAN TROGLODYTES (CHIMPANZEE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
 OC CATARRHINI; HOMINIDAE; PAN.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 94292799.

RA BIRNICHAM D.J., SHEN X.P., HOURCADE D., NICKELIS M.W., ATKINSON J.P.;  
RT "Primary sequence of an alternatively spliced form of CRL. Candidate  
for the 75,000 M(r) complement receptor expressed on chimpanzee  
erythrocytes."  
RL J. IMMUNOL. 153:691-700(1994).  
DR EMBL: L24920; G557725; -  
PFAM: PF00084; sushi; 30.  
FT NON\_TER 1  
SO SEQUENCE 2014 AA; 221280 MW; 1370CDD8 CRC32;

Query Match 95.9%; Score 118; DB 4; Length 2014;  
Best Local Similarity 95.7%; Pred. No. 5.7e-10;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FELVGEPSIYCTSDNDQVGWSG 23  
Db 180 FELVGEPSIYCTSDNDQVGWSG 202

RESULT 5  
O28770 PRELIMINARY; PRT; 315 AA.  
ID 028770;  
AC 028770;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE COMPLEMENT RECEPTOR (FRAGMENT).  
OS PAPIO HAMADRYAS (HAMADRYAS BABOON).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; PAPIO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BONE MARROW;  
RA BIRNICHAM D.J., LOGAR C.M., SHEN X.P., CHEN W.;  
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: L77978; G1301611; -  
DR PFAM: PF00084; sushi; 5.  
FT NON\_TER 1  
FT NON\_TER 1  
SO SEQUENCE 315 AA; 35172 MW; 5776DB8A CRC32;

Query Match 92.7%; Score 114; DB 4; Length 315;  
Best Local Similarity 91.3%; Pred. No. 3e-10;  
Matches 21; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FELVGEPSIYCTSDNDQVGWSG 23  
Db 164 FELVGEPSIYCTSDNDQVGWSG 186

RESULT 6  
O29528 PRELIMINARY; PRT; 1911 AA.  
ID 029528;  
AC 029528;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE COMPLEMENT RECEPTOR 1 (FRAGMENT).  
GN CRL.  
OS PAPIO HAMADRYAS (HAMADRYAS BABOON).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; PAPIO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA CLEMENZA L., SUBRAMANIAN B.V., NICKELIS M.W., HOURCADE D.E.,  
RA ATKINSON J.P.;  
RL SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: L39781; G662829; -  
DR PFAM: PF00084; sushi; 29.  
FT NON\_TER 1911

SO SEQUENCE 1911 AA; 210173 MW; 03D640C2 CRC32;

Query Match 92.7%; Score 114; DB 4; Length 1911;  
Best Local Similarity 91.3%; Pred. No. 2.3e-09;  
Matches 21; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FELVGEPSIYCTSDNDQVGWSG 23  
Db 92 FELVGEPSIYCTSDNDQVGWSG 114

RESULT 7  
O14079 PRELIMINARY; PRT; 479 AA.  
ID 014079;  
AC 014079;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE COMPLEMENT RECEPTOR 1-LIKE PROTEIN (CRL) (FRAGMENT).  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 90110163.  
RA HOURCADE D., MIESNER D.R., BEE C., ZEIDES W., ATKINSON J.P.;  
RT "Duplication and divergence of the amino-terminal coding region of the  
RT complement receptor 1 (CRL) gene. An example of concerted (horizontal)  
evolution within a gene."  
RL J. BIOL. CHEM. 265:974-980(1990).  
DR EMBL: M31231; G563326; JOINED.  
DR EMBL: M31232; G563326; JOINED.  
DR EMBL: M31233; G563326; JOINED.  
DR EMBL: M31234; G563326; JOINED.  
DR EMBL: M31235; G563326; JOINED.  
DR EMBL: M31236; G563326; JOINED.  
DR EMBL: M31237; G563326; JOINED.  
DR EMBL: M31238; G563326; JOINED.  
DR EMBL: M31239; G563326; JOINED.  
DR EMBL: M31240; G563326; JOINED.  
DR EMBL: M31241; G563326; JOINED.  
DR EMBL: M31242; G563326; JOINED.  
DR EMBL: M31243; G563326; JOINED.  
DR EMBL: M31244; G563326; JOINED.  
DR EMBL: M31245; G563326; JOINED.  
DR EMBL: M31246; G563326; JOINED.  
DR PFAM: PF00084; sushi; 7.  
FT NON\_TER 479  
FT NON\_TER 479  
SO SEQUENCE 479 AA; 52533 MW; EC994E0D CRC32;

Query Match 91.1%; Score 112; DB 2; Length 479;  
Best Local Similarity 91.3%; Pred. No. 9.8e-10;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FELVGEPSIYCTSDNDQVGWSG 23  
Db 205 FELVGEPSIYCTSDNDQVGWSG 227

RESULT 8  
O28769 PRELIMINARY; PRT; 522 AA.  
ID 028769;  
AC 028769;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE COMPLEMENT RECEPTOR (FRAGMENT).  
OS PAPIO HAMADRYAS (HAMADRYAS BABOON).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; PAPIO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BONE MARROW;  
RA BIRNICHAM D.J., LOGAR C.M., SHEN X.P., CHEN W.;  
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: L77977; G1301609; -  
DR PFAM: PF00084; sushi; 7.

FT NON\_TER 1 1  
 SQ SEQUENCE 522 AA; 56626 MW; 7862072C CRC32;

Query Match  
 Best Local Similarity 91.1%; Score 112; DB 4; Length 522;  
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FELVGPSTYSTNDQVIGWSG 23  
 Db 210 FELVGPSTYSTNDQVIGWSG 232

RESULT 9  
 ID 028797 PRELIMINARY; PRT; 522 AA.

AC 028797;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DE UNKNOWN PROTEIN (FRAGMENT).  
 OS PAN TROGLODYTES (CHIMPANZEE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
 OC CATARRHINI; HOMINIDAE; PAN.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 94292799.  
 RA BIRMINGHAM D.J., SHEN X.P., HOURCADE D., NICKELLS M.W., ATKINSON J.P.;  
 RT "Primary sequence of an alternatively spliced form of CRL. Candidate  
 for the 75,000 M(r) complement receptor expressed on chimpanzee  
 erythrocytes."  
 RL J. IMMUNOL. 153:691-700(1994).  
 DR EMBL; L24922; G557729; -  
 DR PFAM; PF00084; sushi; 8.  
 FT NON\_TER 1  
 SQ SEQUENCE 522 AA; 57304 MW; 21A32E12 CRC32;

Query Match  
 Best Local Similarity 91.1%; Score 112; DB 4; Length 522;  
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FELVGPSTYSTNDQVIGWSG 23  
 Db 147 FELVGPSTYSTNDQVIGWSG 169

RESULT 10  
 ID 064735 PRELIMINARY; PRT; 483 AA.

AC 064735;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE COMPLEMENT RECEPTOR RELATED PROTEIN.  
 GN CRRY.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C;  
 RX MEDLINE; 90171600.  
 RA PAUL M.S., AEGERTER-SHAW M., CEPER K., MILLER M.D., WEIS J.H.;  
 RT "The murine complement receptor gene family. III. The genomic and  
 transcriptional complexity of the Crry and Crry-ps genes."  
 RL J. IMMUNOL. 144:1988-1996(1990).  
 DR EMBL; M34166; G387133; JOINED.  
 DR EMBL; M34167; G387133; JOINED.  
 DR EMBL; M34168; G387133; JOINED.  
 DR EMBL; M34169; G387133; JOINED.  
 DR EMBL; M34170; G387133; JOINED.  
 DR EMBL; M34171; G387133; JOINED.

DR EMBL; M34172; G387133; JOINED.  
 DR EMBL; M34173; G387133; -  
 DR EMBL; M34174; G387133; -  
 DR EMBL; M34164; G387133; JOINED.  
 DR EMBL; M34165; G387133; JOINED.  
 DR EMBL; M34166; G387133; JOINED.  
 DR EMBL; M34170; G387133; JOINED.  
 DR EMBL; M34171; G387133; JOINED.  
 DR EMBL; M34172; G387133; JOINED.  
 DR EMBL; M34165; G387133; JOINED.  
 DR EMBL; M34166; G387133; JOINED.  
 DR EMBL; M34167; G387133; JOINED.  
 DR EMBL; M34168; G387133; JOINED.  
 DR EMBL; M34169; G387133; JOINED.  
 DR MGD; MGI:88513; CRRY.  
 DR PFAM; PF00084; sushi; 5.  
 FT VARIANT 41  
 SQ SEQUENCE 483 AA; 53762 MW; 70C73A80 CRC32; MISSING (IN LIVER ISOFORM).

Query Match  
 Best Local Similarity 80.5%; Score 99; DB 10; Length 483;  
 Matches 16; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 FELVGPSTYSTNDQVIGWSG 23  
 Db 247 FNLVGPSTYSTNDQVIGWSG 269

RESULT 11  
 ID 035520 PRELIMINARY; PRT; 417 AA.

AC 035520;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE 512 ANTIGEN (FRAGMENT).  
 OS RATTUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LIVER;  
 RA DOHI N., SAKURADA C., NONAKA M., OKADA N., OKADA H.;  
 RL SUBMITTED (NOV-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; D42116; D1023417; -  
 DR PFAM; PF00084; sushi; 6.  
 FT NON\_TER 417  
 SQ SEQUENCE 417 AA; 45951 MW; A6F919B4 CRC32;

Query Match  
 Best Local Similarity 74.0%; Score 91; DB 10; Length 417;  
 Matches 17; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 FELVGPSTYSTNDQVIGWSG 23  
 Db 62 FNLVGPSTYSTNDQVIGWSG 84

RESULT 12  
 ID 063612 PRELIMINARY; PRT; 497 AA.

AC 063612;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE 512 ANTIGEN PRECURSOR.  
 OS RATTUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC TISSUE-LIVER:  
 RX MEDLINE: 94161746.  
 RA SAKURADA C., SENO H., DOHI N., TAKIZAWA H., NONAKA M., OKADA N.,  
 RA OKADA H.;  
 RT "Molecular cloning of the rat complement regulatory protein, 512  
 RT antigen.";  
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 198;819-826(1994).  
 DR EMBL: D42114; G987688; -.  
 DR PFAM: PF00084; sushi; 6.  
 KW SIGNAL.  
 FT SIGNAL.  
 FT CHAIN 37 497 512 ANTIGEN.  
 SQ SEQUENCE 497 AA; 54786 MW; 076547C7 CRC32;

Query Match 74.0%; Score 91; DB 10; Length 497;  
 Best Local Similarity 73.9%; Pred. No. 2e-06;  
 Matches 17; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 FELVGEPSISTSDNOVGWISG 23  
 ID 063135 PRELIMINARY; PRT; 559 AA.  
 AC 063135;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE COMPLEMENT REGULATORY PROTEIN.  
 GN CRRY.  
 OS RATIUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCIUROGNATHI; MURIDAE; MORINAE; RATTUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY;  
 RX MEDLINE: 96006570.  
 RA OUIRG R.J., LO C.F., ALEXANDER J.J., SNEED A.E., MOXLEY G. III;  
 RT "Molecular characterization of rat Crry: widespread distribution of  
 RT two alternative forms of Crry mRNA.";  
 RL IMMUNOGENETICS 42:362-367(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LIVER:  
 RA DOHI N., SAKURADA C., NONAKA M., OKADA N., OKADA H.;  
 RA SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: L36532; G120295; -.  
 DR EMBL: D42115; D1023416; -.  
 DR PFAM: PF00084; sushi; 7.  
 SQ SEQUENCE 559 AA; 61680 MW; C87EEC58 CRC32;

RESULT 13  
 ID 063135 PRELIMINARY; PRT; 559 AA.  
 AC 063135;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE COMPLEMENT REGULATORY PROTEIN.  
 GN CRRY.  
 OS RATIUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCIUROGNATHI; MURIDAE; MORINAE; RATTUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY;  
 RX MEDLINE: 96006570.  
 RA OUIRG R.J., LO C.F., ALEXANDER J.J., SNEED A.E., MOXLEY G. III;  
 RT "Molecular characterization of rat Crry: widespread distribution of  
 RT two alternative forms of Crry mRNA.";  
 RL IMMUNOGENETICS 42:362-367(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LIVER:  
 RA DOHI N., SAKURADA C., NONAKA M., OKADA N., OKADA H.;  
 RA SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: L36532; G120295; -.  
 DR EMBL: D42115; D1023416; -.  
 DR PFAM: PF00084; sushi; 7.  
 SQ SEQUENCE 559 AA; 61680 MW; C87EEC58 CRC32;

Query Match 74.0%; Score 91; DB 10; Length 559;  
 Best Local Similarity 73.9%; Pred. No. 2.3e-06;  
 Matches 17; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 FELVGEPSISTSDNOVGWISG 23  
 ID 063135 PRELIMINARY; PRT; 559 AA.  
 AC 063135;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE COMPLEMENT RECEPTOR TYPE 2 PRECURSOR (CR2).  
 GN (ALTERNATIVELY SPLICED VERSION) (FRAGMENT).  
 DE

RESULT 14

QY 099254 PRELIMINARY; PRT; 679 AA.  
 AC 099254;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE COMPLEMENT RECEPTOR TYPE 2 PRECURSOR (CR2).  
 GN (ALTERNATIVELY SPLICED VERSION) (FRAGMENT).  
 DE

QY 099254 PRELIMINARY; PRT; 679 AA.  
 AC 099254;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE COMPLEMENT RECEPTOR TYPE 2 PRECURSOR (CR2).  
 GN (ALTERNATIVELY SPLICED VERSION) (FRAGMENT).  
 DE

GN CR2.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCIUROGNATHI; MURIDAE; MORINAE; MUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C;  
 RX MEDLINE: 90229754.  
 RA KURTZ C.B., O'TOOLE E., CHRISTENSEN S.M., WEIS J.H.;  
 RT "The murine complement receptor gene family. IV. Alternative splicing  
 RT of Cr2 gene transcripts predicts two distinct gene products that share  
 RT homologous domains with both human CR2 and CRI.";  
 RL J. IMMUNOL. 144:3581-3591(1990).  
 RN [2]  
 RP SEQUENCE OF 21-367 FROM N.A.  
 RX MEDLINE: 95105691.  
 RA KIM Y.U., KINOSHITA T., MOLINA H., HOURCADE D., SEVA T., WAGNER L.M.,  
 RA HOLERS V.M.;  
 RT "Mouse complement regulatory protein Crry/p65 uses the specific  
 RT mechanisms of both human decay-accelerating factor and membrane  
 RT cofactor protein.";  
 RL J. EXP. MED. 181:151-159(1995).  
 CC -1- TISSUE SPECIFICITY: B LYMPHOCYTES.  
 CC -1- ALTERNATIVE PRODUCTS: ALTERNATIVE SPLICING OF CR2 GENE  
 CC TRANSCRIPTS PREDICTS TWO DISTINCT GENE PRODUCTS THAT SHARE  
 CC HOMOLOGOUS DOMAINS WITH BOTH HUMAN CR2 AND CRI.  
 DR EMBL: U17124; G595982; JOINED.  
 DR EMBL: U17128; G595982; JOINED.  
 DR EMBL: U17123; G595982; JOINED.  
 DR EMBL: M36470; G192691; -.  
 DR EMBL: U17127; G595982; JOINED.  
 DR EMBL: U17126; G595982; JOINED.  
 DR EMBL: U17125; G595982; JOINED.  
 DR MGI: MGI:88489; CR2.  
 DR PFAM: PF00084; sushi; 10.  
 KW COMPLEMENT PATHWAY; GLYCOPROTEIN; TRANSMEMBRANE; REPEAT; SIGNAL;  
 KM ALTERNATIVE SPLICING; RECEPTOR.  
 FT SIGNAL 1 11  
 FT CHAIN 12 679 COMPLEMENT RECEPTOR TYPE 2.  
 FT NON\_TER 679 679  
 SQ SEQUENCE 679 AA; 74916 MW; BF08AC75 CRC32;

Query Match 70.7%; Score 87; DB 10; Length 679;  
 Best Local Similarity 72.7%; Pred. No. 1.2e-05;  
 Matches 16; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 FELVGEPSISTSDNOVGWISG 22  
 ID 063129 PRELIMINARY; PRT; 89 AA.  
 AC 063129;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE COMPLEMENT RECEPTOR TYPE 1 (FRAGMENT).  
 GN CRI.  
 OS RATIUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCIUROGNATHI; MURIDAE; MORINAE; RATTUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY;  
 RX MEDLINE: 95002704.  
 RA OUIRG R.J., SNEED A.E.;  
 RT "Molecular characterization of rat glomerular epithelial cell  
 RT complement receptors.";  
 RL J. AM. SOC. NEPHROL. 4:1912-1919(1994).  
 DR EMBL: L19118; G682650; -.

RESULT 15

QY 063129 PRELIMINARY; PRT; 89 AA.  
 AC 063129;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE COMPLEMENT RECEPTOR TYPE 1 (FRAGMENT).  
 GN CRI.  
 OS RATIUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCIUROGNATHI; MURIDAE; MORINAE; RATTUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY;  
 RX MEDLINE: 95002704.  
 RA OUIRG R.J., SNEED A.E.;  
 RT "Molecular characterization of rat glomerular epithelial cell  
 RT complement receptors.";  
 RL J. AM. SOC. NEPHROL. 4:1912-1919(1994).  
 DR EMBL: L19118; G682650; -.

QY 063129 PRELIMINARY; PRT; 89 AA.  
 AC 063129;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE COMPLEMENT RECEPTOR TYPE 1 (FRAGMENT).  
 GN CRI.  
 OS RATIUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCIUROGNATHI; MURIDAE; MORINAE; RATTUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY;  
 RX MEDLINE: 95002704.  
 RA OUIRG R.J., SNEED A.E.;  
 RT "Molecular characterization of rat glomerular epithelial cell  
 RT complement receptors.";  
 RL J. AM. SOC. NEPHROL. 4:1912-1919(1994).  
 DR EMBL: L19118; G682650; -.

